

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2001, 12:17:57 ; Search time 6270.82 Seconds
(without alignments)
11955.722 Million cell updates/sec

Title: US-09-528-031-1
Perfect score: 4847
Sequence: 1 GCGTCATGCTCGGAGCGTG.....AAAAAAAAAGGCGGCCGC 4847

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rtd: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rtd: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_v1: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v1: *
59: gb_v12: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_ro1: *
95: gb_ro2: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query Length | ID | Description |
|------------|--------|-------|--------------|----|----------------------|
| 1 | 4847 | 100.0 | 4847 | 9 | AR099619 Sequence |
| 2 | 4792.6 | 98.9 | 5838 | 88 | AF104942 Homo sapi |
| 3 | 4789.6 | 98.8 | 5881 | 89 | AF146074 Homo sapi |
| 4 | 4788 | 98.8 | 5926 | 97 | HS083661 Homo sapien |
| 5 | 4775.4 | 98.5 | 4781 | 9 | AR091520 Sequence |
| 6 | 4728 | 97.5 | 5728 | 85 | AB019002 Homo sapi |
| 7 | 3774 | 77.9 | 5400 | 94 | AB019003 Mus muscu |
| 8 | 3703.8 | 76.4 | 4939 | 85 | AB005659 Homo sapi |

| | | | | | | |
|----|--------|------|--------|----|-----------|-------------------|
| 9 | 3691.4 | 36.2 | 5744 | 94 | AB020209 | Rattus norvegicus |
| 10 | 1674.4 | 34.5 | 1761 | 9 | A64695 | Sequence 61 |
| 11 | 1054.2 | 21.7 | 2167 | 9 | A64696 | Sequence 62 |
| 12 | 776 | 16.0 | 4492 | 89 | AF352562 | Homo sapiens |
| 13 | 545 | 11.2 | 765 | 94 | AB012090 | Mus musculus |
| 14 | 534.2 | 11.0 | 1076 | 94 | AF213387 | Mus musculus |
| 15 | 511.2 | 10.5 | 5011 | 9 | AR070199 | Mus musculus |
| 16 | 511.2 | 10.5 | 5011 | 9 | AR093440 | Sequence |
| 17 | 511.2 | 10.5 | 5011 | 97 | HMMRFX | Sequence |
| 18 | 508 | 10.5 | 5867 | 97 | SATHRMP | Sequence |
| 19 | 506.4 | 10.4 | 5011 | 9 | AR070198 | Sequence |
| 20 | 506.4 | 10.4 | 5011 | 9 | AR093439 | Sequence |
| 21 | 506.4 | 10.4 | 5011 | 10 | 117455 | Sequence 1 |
| 22 | 479.8 | 9.9 | 188429 | 73 | AC068644 | Homo sapiens |
| 23 | 437.4 | 9.0 | 2940 | 88 | AF071203 | Homo sapiens |
| 24 | 437.4 | 9.0 | 4231 | 88 | AF071202 | Homo sapiens |
| 25 | 420.4 | 8.7 | 4587 | 94 | AF022908 | Mus musculus |
| 26 | 420.4 | 8.7 | 5889 | 9 | AR070200 | Sequence |
| 27 | 420.4 | 8.7 | 5889 | 9 | AR093441 | Sequence |
| 28 | 406.6 | 8.4 | 2969 | 95 | RN0277881 | Sequence |
| 29 | 390.4 | 8.1 | 4880 | 94 | D68086 | Rattus norvegicus |
| 30 | 390.4 | 8.1 | 4918 | 9 | A64636 | Sequence 2 |
| 31 | 390.4 | 8.1 | 4918 | 95 | RATCMOAT | Rattus norvegicus |
| 32 | 388.8 | 8.0 | 4909 | 95 | RNCRMP | Rattus norvegicus |
| 33 | 385.4 | 8.0 | 5023 | 89 | AK000002 | Homo sapiens |
| 34 | 384 | 7.9 | 5407 | 7 | OC24914 | Homo sapiens |
| 35 | 383.8 | 7.9 | 4868 | 97 | HS063970 | Human canal |
| 36 | 382.2 | 7.9 | 5300 | 97 | HS049248 | Human canal |
| 37 | 382.2 | 7.9 | 5586 | 9 | A64635 | Sequence 1 |
| 38 | 380.6 | 7.9 | 4864 | 93 | HSCMRP | H. sapiens m |
| 39 | 380.6 | 7.9 | 4868 | 10 | E15807 | Human mRNA |
| 40 | 376.8 | 7.8 | 4749 | 95 | RNBLSTFLR | R. norvegicus |
| 41 | 376.8 | 7.8 | 4780 | 94 | AB052294 | Rattus norvegicus |
| 42 | 376.8 | 7.8 | 5099 | 95 | RAT5UR | Rattus norvegicus |
| 43 | 375.2 | 7.7 | 166918 | 4 | AC009205 | Drosophila |
| 44 | 375.2 | 7.7 | 170546 | 4 | AC011756 | Drosophila |
| 45 | 375.2 | 7.7 | 265010 | 5 | AE003661 | Drosophila |

ALIGNMENTS

AB020209 Rattus 61
A64656 Sequence 62
AF352582 Homo sap
AB021090 Mus musc
AF213187 Mus musc
AF070199 Sequence
AR093440 Sequence
L05528 Human mult
X78338 Synthetic i
AR070198 Sequence
AR093439 Sequence
I17455 Sequence 1
AC086644 Homo sap
AF081203 Homo sap
AF071202 Homo sap
AF022908 Mus musc
AR070200 Sequence
AR093441 Sequence
AJ277881 Rattus n
D86086 Rat mRNA f
A64656 Sequence 2
L69397 Rattus nor
X93939 Rattus nor
AK000002 Homo sap
Z49144 O. cuniculus
U63370 Homo cana
U64948 Human cana
A64655 Sequence 1
A64655 Sequence 1
E15807 H.sapiens
X97379 R.norvegicus
AB052294 Rattus n
L40624 Rattus nor
AC009205 Drosophila
AF011756 Drosophila
AE003661 Drosophila

[illegible]

| | | | |
|----|------|--|------|
| Db | 61 | GAATTCGATGTGTAACCTAACAAGTCTGTGAGGCCCTTGAAACCTCCACTCAGAGAAAGATGAA | 120 |
| Qy | 121 | GGATATGCAATAGGAAAAAGATATATCATTCCTCCAGTCTCTGGGTATTAAGAAGTGAAGGA | 180 |
| Db | 121 | GGATATGCAATAGGAAAAAGATATATCATTCCTCCAGTCTCTGGGTATTAAGAAGTGAAGGA | 180 |
| Qy | 181 | GAGAAACGACATCTTGGGGACGACAGAGACCGGTGAAGATTCACAGTTCCAGAGAAATCG | 240 |
| Db | 181 | GAGAAACGACATCTTGGGGACGACAGAGACCGGTGAAGATTCACAGTTCCAGAGAAATCG | 240 |
| Qy | 241 | ACCGTTGGAAATGCCAAGATATGCTTTGGAAACAGACGCCGAGGCCGAGGGGCTCTCTTGA | 300 |
| Db | 241 | ACCGTTGGAAATGCCAAGATATGCTTTGGAAACAGACGCCGAGGCCGAGGGGCTCTCTTGA | 300 |
| Qy | 301 | TGCTTCATGATTTCTACAGTCTGCAATTCCTGGATGAGAGACATCCCAAGGAAAGTACCA | 360 |
| Db | 301 | TGCTTCATGATTTCTACAGTCTGCAATTCCTGGATGAGAGACATCCCAAGGAAAGTACCA | 360 |
| Qy | 361 | TCATGGCTTAGAGTCTGTAAGCCCATCCGGAGTACTTCACAAACACAGACCCAGTGA | 420 |
| Db | 361 | TCATGGCTTAGAGTCTGTAAGCCCATCCGGAGTACTTCACAAACACAGACCCAGTGA | 420 |
| Qy | 421 | CAATGCTGGGCTTTTTCCTGATATGACTTTTTCGTTCTCTGACCCGTGTGGC | 480 |
| Db | 421 | CAATGCTGGGCTTTTTCCTGATATGACTTTTTCGTTCTCTGACCCGTGTGGC | 480 |
| Qy | 481 | CCACAAAGAGGGGAGCTCTCAATGGAAACAGTGTGTCTGTGTCCAGACAGATCTTC | 540 |
| Db | 481 | CCACAAAGAGGGGAGCTCTCAATGGAAACAGTGTGTCTGTGTCCAGACAGATCTTC | 540 |
| Qy | 541 | TGACGTAACTGAGAGAAATAGAGAGCTGTGGCAAGAGAGCTGAATGAATGGGACC | 600 |
| Db | 541 | TGACGTAACTGAGAGAGACTAGAGAGACTGTGTGGCAAGAGAGCTGAATGAATGGGACC | 600 |
| Qy | 601 | AGAGCGTCTTCCCTGGGAAAGGGTTGTGTGATCTTCTGCCAGACAGGCTCATCTGC | 660 |
| Db | 601 | AGAGCGTCTTCCCTGGGAAAGGGTTGTGTGATCTTCTGCCAGACAGGCTCATCTGC | 660 |
| Qy | 661 | CATGCTGTGCTTATGATATCAGCAGCTGGCTGCTCAAGTGAACAGCTTATGTGTAA | 720 |
| Db | 661 | CATGCTGTGCTTATGATATCAGCAGCTGGCTGCTCAAGTGAACAGCTTATGTGTAA | 720 |
| Qy | 721 | ACACCTCTTGGAGATATACCCAGGCAACAGATCTAACCTGACAGTCTGTGTGTAGT | 780 |
| Db | 721 | ACACCTCTTGGAGATATACCCAGGCAACAGATCTAACCTGACAGTCTGTGTGTAGT | 780 |
| Qy | 781 | GCATGGGCTCTCTCTGACGGAATCGTGGGCTTGTGGTGCATGACTTGGGCAAT | 840 |
| Db | 781 | GCATGGGCTCTCTCTGACGGAATCGTGGGCTTGTGGTGCATGACTTGGGCAAT | 840 |
| Qy | 841 | GAATTAACCGAACGGGTGCCGCTTGGGGGGGGCCATCTTAACCATGGCATTTTAAGAAGAT | 900 |
| Db | 841 | GAATTAACCGAACGGGTGCCGCTTGGGGGGGGCCATCTTAACCATGGCATTTTAAGAAGAT | 900 |
| Qy | 901 | CTTTAAATTTAAAGAACTTTAAAGAAATTCCTGGGTGAGCTCATCAACATTTGCTCCA | 960 |
| Db | 901 | CTTTAAATTTAAAGAACTTTAAAGAAATTCCTGGGTGAGCTCATCAACATTTGCTCCA | 960 |
| Qy | 961 | CGATGGGCAAGATGTTTGAAGCAGACAGCCGTTGGCAGCTGTGTGGTGGAGGACCCGT | 1020 |
| Db | 961 | CGATGGGCAAGATGTTTGAAGCAGACAGCCGTTGGCAGCTGTGTGGTGGAGGACCCGT | 1020 |
| Qy | 1021 | TGTTGCAATCTTAGCGATGATTTTATATGATTAATTTCTGGGACCAACAGGCTTCTGGG | 1080 |
| Db | 1021 | TGTTGCAATCTTAGCGATGATTTTATATGATTAATTTCTGGGACCAACAGGCTTCTGGG | 1080 |
| Qy | 1081 | ATCAGCTGTTTATCTCTTTTAAACCAAGATGATGTTTGCATCAAGGCTCACAGCAT | 1140 |
| Db | 1081 | ATCAGCTGTTTATCTCTTTTAAACCAAGATGATGTTTGCATCAAGGCTCACAGCAT | 1140 |
| Qy | 1141 | TTTTCAGAGAAATGCGTGGCCGCCACAGATACACTGTCCAGAGATGAATGAAGTCT | 1200 |
| Db | 1141 | TTTTCAGAGAAATGCGTGGCCGCCACAGATGAACGTGTCCAGAGATGAATGAAGTCT | 1200 |

1201 TACTTACATTAATTAATCAAAATGATGCGCTGGGGCAAAAGCATTTTCTCAGAGTGTCA 1260
1201 TACTTACATTAATTAATCAAAATGATGCGCTGGGGCAAAAGCATTTTCTCAGAGTGTCA 1260
1261 GAAATCCGCGAGAGAGAGCGCTGGATATTGGAAAAAGCGGGTACTTCTCAGAGCATCAC 1320
1261 GAAATCCGCGAGAGAGAGCGCTGGATATTGGAAAAAGCGGGTACTTCTCAGAGCATCAC 1320
1321 TGTGGGTGTGGCTCCCATTTGTGTGTGATTTGCCAGCGTGTGTGACTTCTCTGTTCATAT 1380
1321 TGTGGGTGTGGCTCCCATTTGTGTGTGATTTGCCAGCGTGTGTGACTTCTCTGTTCATAT 1380
1321 TGTGGGTGTGGCTCCCATTTGTGTGTGATTTGCCAGCGTGTGTGACTTCTCTGTTCATAT 1380
1381 GACCGTGGCTTCGATCTGACAGCAGCAGAGCTTTACAGAGTGTGACAGTCTTCAATTC 1440
1381 GACCGTGGCTTCGATCTGACAGCAGCAGAGCTTTACAGAGTGTGACAGTCTTCAATTC 1440
1381 GACCGTGGCTTCGATCTGACAGCAGCAGAGCTTTACAGAGTGTGACAGTCTTCAATTC 1440
1441 CATGACTTTTGTCTTGAAGATTAACACCGTTTTCAGTAAAGTCCCTCTCAGAAAGCTCAGT 1500
1441 CATGACTTTTGTCTTGAAGATTAACACCGTTTTCAGTAAAGTCCCTCTCAGAAAGCTCAGT 1500
1501 GCGTGTGACAAATTTAGAGTGTGTCTTAATGGAAGGTTTACATATATAAGACAA 1560
1501 GCGTGTGACAAATTTAGAGTGTGTCTTAATGGAAGGTTTACATATATAAGACAA 1560
1501 GCGTGTGACAAATTTAGAGTGTGTCTTAATGGAAGGTTTACATATATAAGACAA 1560
1561 ACCAGCAGATCCTCAGATCAATGATAGATGAAAAATGACACCTTGGCATGGGACTCTTC 1620
1561 ACCAGCAGATCCTCAGATCAATGATAGATGAAAAATGACACCTTGGCATGGGACTCTTC 1620
1561 ACCAGCAGATCCTCAGATCAATGATAGATGAAAAATGACACCTTGGCATGGGACTCTTC 1620
1621 CCATCCAGATTCAGAAAGTCCGCCAAGCTGACCCCAAAATGAAAAAGACAAAGAGGCG 1680
1621 CCATCCAGATTCAGAAAGTCCGCCAAGCTGACCCCAAAATGAAAAAGACAAAGAGGCG 1680
1621 CCATCCAGATTCAGAAAGTCCGCCAAGCTGACCCCAAAATGAAAAAGACAAAGAGGCG 1680
1681 TTCAGAGGCGCAAGAAAGAGAGAGTGTGAGCTGACAGCGCTGAGCATCAGAGCGGTGCT 1740
1681 TTCAGAGGCGCAAGAAAGAGAGAGTGTGAGCTGACAGCGCTGAGCATCAGAGCGGTGCT 1740
1681 TTCAGAGGCGCAAGAAAGAGAGAGTGTGAGCTGACAGCGCTGAGCATCAGAGCGGTGCT 1740
1741 GGCAGAGCGAAAGAGGCGACCTCTCTCTGAGACAGTGAAGCGGCGCCAGTCCGAAAGAGA 1800
1741 GGCAGAGCGAAAGAGGCGACCTCTCTCTGAGACAGTGAAGCGGCGCCAGTCCGAAAGAGA 1800
1741 GGCAGAGCGAAAGAGGCGACCTCTCTCTGAGACAGTGAAGCGGCGCCAGTCCGAAAGAGA 1800
1801 AGAAGGCAAGCAGATCCACCTGGGCGACCTGGCTTACAGAGGACACTGACAGCATCGA 1860
1801 AGAAGGCAAGCAGATCCACCTGGGCGACCTGGCTTACAGAGGACACTGACAGCATCGA 1860
1801 AGAAGGCAAGCAGATCCACCTGGGCGACCTGGCTTACAGAGGACACTGACAGCATCGA 1860
1861 TGTGAGATCCAAAGAGGTAATCTGTTGGAATCTGCGCAGTGTGGAAAGTGGAAAAAC 1920
1861 TGTGAGATCCAAAGAGGTAATCTGTTGGAATCTGCGCAGTGTGGAAAGTGGAAAAAC 1920
1861 TGTGAGATCCAAAGAGGTAATCTGTTGGAATCTGCGCAGTGTGGAAAGTGGAAAAAC 1920
1921 CTTCTCATTTTCCAGCATTTTAGGCGAGATGACGCTTCTAGAGGCGAGCATTTGCATTCAG 1980
1921 CTTCTCATTTTCCAGCATTTTAGGCGAGATGACGCTTCTAGAGGCGAGCATTTGCATTCAG 1980
1921 CTTCTCATTTTCCAGCATTTTAGGCGAGATGACGCTTCTAGAGGCGAGCATTTGCATTCAG 1980
1981 TGGAACTTTCGCTTATGTGTGGCCAGAGCGCTGATCTCAATGTACTCTGAGAGACAA 2040
1981 TGGAACTTTCGCTTATGTGTGGCCAGAGCGCTGATCTCAATGTACTCTGAGAGACAA 2040
1981 TGGAACTTTCGCTTATGTGTGGCCAGAGCGCTGATCTCAATGTACTCTGAGAGACAA 2040
2041 CATCTGTTTGGGAAGAAATATGATGAAGAAAGATCAACTCTGTGTCTGAACAGCTGTG 2100
2041 CATCTGTTTGGGAAGAAATATGATGAAGAAAGATCAACTCTGTGTCTGAACAGCTGTG 2100
2041 CATCTGTTTGGGAAGAAATATGATGAAGAAAGATCAACTCTGTGTCTGAACAGCTGTG 2100
2101 CCGTGGAGCTGACGCTGGCCATTTCTTCCAGACGCGCTGACGAGGAATTTGGAGAGGAG 2160
2101 CCGTGGAGCTGACGCTGGCCATTTCTTCCAGACGCGCTGACGAGGAATTTGGAGAGGAG 2160
2101 CCGTGGAGCTGACGCTGGCCATTTCTTCCAGACGCGCTGACGAGGAATTTGGAGAGGAG 2160
2161 AGCAACCTGAGAGCGGTGGGCGACGCGAGAGATCAAGCTTGTGCGCGGCTGTGTATGTGA 2220
2161 AGCAACCTGAGAGCGGTGGGCGACGCGAGAGATCAAGCTTGTGCGCGGCTGTGTATGTGA 2220
2161 AGCAACCTGAGAGCGGTGGGCGACGCGAGAGATCAAGCTTGTGCGCGGCTGTGTATGTGA 2220
2221 CAGAGCATCTACATCTGAGAGACCCCTCAGTGTCTTATGATGCTCATGTGGGCAACCA 2280
2221 CAGAGCATCTACATCTGAGAGACCCCTCAGTGTCTTATGATGCTCATGTGGGCAACCA 2280

2281 CATCTTCAATAGTGCATTCGCGAAGCATCTCAAGTCCAGAGCTGTCTGTACCA 2340
2281 CATCTTCAATAGTGCATTCGCGAAGCATCTCAAGTCCAGAGCTGTCTGTACCA 2340
2281 CATCTTCAATAGTGCATTCGCGAAGCATCTCAAGTCCAGAGCTGTCTGTACCA 2340
2341 CCAGTTACAGTACTGTGTGACTGTGATGATGATCTTCAATGAAGAGGCTGTATTAC 2400
2341 CCAGTTACAGTACTGTGTGACTGTGATGATGATCTTCAATGAAGAGGCTGTATTAC 2400
2341 CCAGTTACAGTACTGTGTGACTGTGATGATGATCTTCAATGAAGAGGCTGTATTAC 2400
2401 GGAAGAGGCGACCCATGAGAGCATGATTAATGATGATTAATGATGATGATGATGATGAT 2460
2401 GGAAGAGGCGACCCATGAGAGCATGATTAATGATGATTAATGATGATGATGATGATGAT 2460
2401 GGAAGAGGCGACCCATGAGAGCATGATTAATGATGATTAATGATGATGATGATGATGAT 2460
2461 TAACCTGTGTGTGGAGAGACACCGCATGATGATGATGATGATGATGATGATGATGAT 2520
2461 TAACCTGTGTGTGGAGAGACACCGCATGATGATGATGATGATGATGATGATGATGAT 2520
2461 TAACCTGTGTGTGGAGAGACACCGCATGATGATGATGATGATGATGATGATGATGAT 2520
2521 TTCACAGAAAGTCAACAGACAGAGGCTCTTAACAGATCAATTAAGAGAAAAAC 2580
2521 TTCACAGAAAGTCAACAGACAGAGGCTCTTAACAGATCAATTAAGAGAAAAAC 2580
2521 TTCACAGAAAGTCAACAGACAGAGGCTCTTAACAGATCAATTAAGAGAAAAAC 2580
2581 AGTAAAGCCAGAGAGAGGCGACCTGTGACGTGGAAGAGAAAGGCGAGGTTCAAGTGC 2640
2581 AGTAAAGCCAGAGAGAGGCGACCTGTGACGTGGAAGAGAAAGGCGAGGTTCAAGTGC 2640
2581 AGTAAAGCCAGAGAGAGGCGACCTGTGACGTGGAAGAGAAAGGCGAGGTTCAAGTGC 2640
2641 CTGTCAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
2641 CTGTCAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
2641 CTGTCAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2700
2701 TATGCGCTTTTTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
2701 TATGCGCTTTTTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
2701 TATGCGCTTTTTCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
2761 CTGATCAAGAGAGAGAGGCGACACCTGTGATGATGATGATGATGATGATGATGATGAT 2820
2761 CTGATCAAGAGAGAGAGGCGACACCTGTGATGATGATGATGATGATGATGATGATGAT 2820
2761 CTGATCAAGAGAGAGAGGCGACACCTGTGATGATGATGATGATGATGATGATGATGAT 2820
2821 TGACAGCATGAAGAGCAATCTCATATGATGATGATGATGATGATGATGATGATGAT 2880
2821 TGACAGCATGAAGAGCAATCTCATATGATGATGATGATGATGATGATGATGATGAT 2880
2821 TGACAGCATGAAGAGCAATCTCATATGATGATGATGATGATGATGATGATGATGAT 2880
2881 GGCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
2881 GGCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
2881 GGCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
2881 GGCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
2941 AGCTTCTCTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
2941 AGCTTCTCTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
2941 AGCTTCTCTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
3001 TTTTGCAG 3060
3001 TTTTGCAG 3060
3001 TTTTGCAG 3060
3001 TTTTGCAG 3060
3061 TGACGTGCGGCTGCGCTTCCAGGCGAGATGATGATGATGATGATGATGATGATGAT 3120
3061 TGACGTGCGGCTGCGCTTCCAGGCGAGATGATGATGATGATGATGATGATGATGAT 3120
3061 TGACGTGCGGCTGCGCTTCCAGGCGAGATGATGATGATGATGATGATGATGATGAT 3120
3061 TGACGTGCGGCTGCGCTTCCAGGCGAGATGATGATGATGATGATGATGATGATGAT 3120
3121 CTGTGTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
3121 CTGTGTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
3121 CTGTGTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
3121 CTGTGTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
3181 CATCTCTTTTTCAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
3181 CATCTCTTTTTCAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
3181 CATCTCTTTTTCAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
3181 CATCTCTTTTTCAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
3241 GGACATATACAGAGAGTCACTTCTCTCCAGATCAAGCTCAGAGATACAGAGGCTTGC 3300
3241 GGACATATACAGAGAGTCACTTCTCTCCAGATCAAGCTCAGAGATACAGAGGCTTGC 3300
3241 GGACATATACAGAGAGTCACTTCTCTCCAGATCAAGCTCAGAGATACAGAGGCTTGC 3300
3241 GGACATATACAGAGAGTCACTTCTCTCCAGATCAAGCTCAGAGATACAGAGGCTTGC 3300
3301 CACATCCAGCGCTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
3301 CACATCCAGCGCTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
3301 CACATCCAGCGCTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
3301 CACATCCAGCGCTTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
3361 TGACAAACCAAGCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 3420
3361 TGACAAACCAAGCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 3420

| | | | |
|----|------|--|------|
| Db | 3361 | TGACAAACAAGCTCCCTTTTTTTTTTTGGTTTACGTGTGGCATGGCGTGGCTGGCTCGGGCT | 3420 |
| Qy | 3421 | GGACCTCATAGCATGGCCCTCATACACACACAGGGGGTGAATGATCGTTCTTATACACGG | 3480 |
| Db | 3421 | GGACCTCATAGCATGGCCCTCATACACACACAGGGGGTGAATGATCGTTCTTATACACGG | 3480 |
| Qy | 3481 | GCAGATTTCCCCACAGCCCTATCGGGGTCTCGGCATCTTTATGCTGCAGTTAAACGGGGCT | 3540 |
| Db | 3481 | GCAGATTTCCCCACAGCCCTATCGGGGTCTCGGCATCTTTATGCTGCAGTTAAACGGGGCT | 3540 |
| Qy | 3541 | GTTCACAGTTAACGGTACAGACTGGCATCTGAGACAGAACTCGATTCACCTCGGTGGAGAG | 3600 |
| Db | 3541 | GTTCACAGTTAACGGTACAGACTGGCATCTGAGACAGAACTCGATTCACCTCGGTGGAGAG | 3600 |
| Qy | 3601 | GATCAATCAGTACATTAAAGACTGTGTCCTTGGAAAGACCTGGCAGAAATTAAACAAGGC | 3660 |
| Db | 3601 | GATCAATCAGTACATTAAAGACTGTGTCCTTGGAAAGACCTGGCAGAAATTAAACAAGGC | 3660 |
| Qy | 3661 | TCCCTCCCTGACTGGCCCCAGAGAGGAGGTGACTCTTTGAAACGACAGATAGAGTA | 3720 |
| Db | 3661 | TCCCTCCCTGACTGGCCCCAGAGAGGAGGTGACTCTTTGAAACGACAGATAGAGTA | 3720 |
| Qy | 3721 | CCGAGAAACCTCCCTCGCTGCTTAAAGAAAGTATCCTTCAGATCAAACTTAAGAGAGA | 3780 |
| Db | 3721 | CCGAGAAACCTCCCTCGCTGCTTAAAGAAAGTATCCTTCAGATCAAACTTAAGAGAGA | 3780 |
| Qy | 3781 | GATTGGCAATTGTGGGGCGACAGAGATCAGGGAAGTCCCTGCTGGGGATGGCCCTCTCCG | 3840 |
| Db | 3781 | GATTGGCAATTGTGGGGCGACAGAGATCAGGGAAGTCCCTGCTGGGGATGGCCCTCTCCG | 3840 |
| Qy | 3841 | TCTGGTGAAGTATCTGAGAGCGTGCATCAAGATTTGATGAGAGTAACTGATGATTTTGG | 3900 |
| Db | 3841 | TCTGGTGAAGTATCTGAGAGCGTGCATCAAGATTTGATGAGAGTAACTGATGATTTTGG | 3900 |
| Qy | 3901 | CCTTGCAGACTCCGAAGCAAACTCTCTATCATCTCCTCAAGAGCCGGTCTCTTCAAGTGG | 3960 |
| Db | 3901 | CCTTGCAGACTCCGAAGCAAACTCTCTATCATCTCCTCAAGAGCCGGTCTCTTCAAGTGG | 3960 |
| Qy | 3961 | CACCTGTCAAGTCAAAATTTGGACCCCTTCAACCGTACGTACATGGAAGACCAATTTGGGATGC | 4020 |
| Db | 3961 | CACCTGTCAAGTCAAAATTTGGACCCCTTCAACCGTACGTACATGGAAGACCAATTTGGGATGC | 4020 |
| Qy | 4021 | CCTGAGAGGACACACATGGAAGAATGTATGTGCAGCTACACCTCAAACTTGAAATCTGA | 4080 |
| Db | 4021 | CCTGAGAGGACACACATGGAAGAATGTATGTGCAGCTACACCTCAAACTTGAAATCTGA | 4080 |
| Qy | 4081 | AGTGATGGAGATGGGGATTACTTCTCAGTGGGGGAACGGCAGCTCTTGTGCATAGCTAG | 4140 |
| Db | 4081 | AGTGATGGAGATGGGGATTACTTCTCAGTGGGGGAACGGCAGCTCTTGTGCATAGCTAG | 4140 |
| Qy | 4141 | AGCCCTCTCCGCCACGTGAAGATTTGATTTAGATGAAGCCACAGCTGCCATGAGACAC | 4200 |
| Db | 4141 | AGCCCTCTCCGCCACGTGAAGATTTGATTTAGATGAAGCCACAGCTGCCATGAGACAC | 4200 |
| Qy | 4201 | AGAGACAGACTTATTTGATTAAGAGACATCCGAGAGAAATTTGCAGACTGTACCATGCT | 4260 |
| Db | 4201 | AGAGACAGACTTATTTGATTAAGAGACATCCGAGAGAAATTTGCAGACTGTACCATGCT | 4260 |
| Qy | 4261 | GACCATTTGCCATCGCTTCGACACAGGTTTACGCTCCGATAGATTAATGTCTGGCCCA | 4320 |
| Db | 4261 | GACCATTTGCCATCGCTTCGACACAGGTTTACGCTCCGATAGATTAATGTCTGGCCCA | 4320 |
| Qy | 4321 | GGGACAGTGGTGAAGTTTACACCCCATGGGTCCTTGTCCAAAGAGATTTCCCGAAT | 4380 |
| Db | 4321 | GGGACAGTGGTGAAGTTTACACCCCATGGGTCCTTGTCCAAAGAGATTTCCCGAAT | 4380 |
| Qy | 4381 | CTATGCCATGTTTGTGCTCTCAGAGAAACAAGTGTGCTCAAGGGGTGACTCCTCCGT | 4440 |
| Db | 4381 | CTATGCCATGTTTGTGCTCTCAGAGAAACAAGTGTGCTCAAGGGGTGACTCCTCCGT | 4440 |
| Qy | 4441 | TGAGGAAGTCTCTTTTCTTTAGAGCATTGCATTCCTCGCTGGGGCGGGCCCTTCATC | 4500 |
| Db | 4441 | TGAGGAAGTCTCTTTTCTTTAGAGCATTGCATTCCTCGCTGGGGCGGGCCCTTCATC | 4500 |

| | | | |
|------------|--|---|----------------------|
| Db | 4441 | TGAGCAAGCTCTCTTTCTTTAGAGCATTCGCCATTTCCGCTGCGGGGCGGCGCCCTTCATC | 4500 |
| OY | 4501 | GGGTCCTCCFACCGAAACCTTGCCCTTCGATTTATCTTTCCGACAGCAGTTCCGGAT | 4560 |
| Db | 4501 | GGGTCCTCCFACCGAAACCTTGCCCTTCGATTTATCTTTCCGACAGCAGTTCCGGAT | 4560 |
| OY | 4561 | TGGCTTGTGTCTCCTTTTAGGAGAGATCATTTGATTATTTATTTCCATAT | 4620 |
| Db | 4561 | TGGCTTGTGTCTCCTTTTAGGAGAGATCATTTGATTATTTATTTCCATAT | 4620 |
| OY | 4621 | TCAGTAAACAAAATTTAGTTTGTCTTTATTTGCACTTAAAGGTTCAGGGAACCGT | 4680 |
| Db | 4621 | TCAGTAAACAAAATTTAGTTTGTCTTTATTTGCACTTAAAGGTTCAGGGAACCGT | 4680 |
| OY | 4681 | TATTTATTTATTCAGAGGCGCTTAAATGAAGCTTTATACGTAATATATTTATATAT | 4740 |
| Db | 4681 | TATTTATTTATTCAGAGGCGCTTAAATGAAGCTTTATACGTAATATATTTATATAT | 4740 |
| OY | 4741 | AATTCGTGACATAGCCCTTATTTACAGTGAATAATTAAGCTGTTATTTATTTAAAT | 4800 |
| Db | 4741 | AATTCGTGACATAGCCCTTATTTACAGTGAATAATTAAGCTGTTATTTATTTAAAT | 4800 |
| OY | 4801 | AAGCATGCTGCTAAAAAAGGCGGCGCGC | 4847 |
| Db | 4801 | AAGCATGCTGCTAAAAAAGGCGGCGCGC | 4847 |
| RESULT | 2 | | |
| LOCUS | AF104942 | 5838 bp | mRNA PRI 12-JAN-1999 |
| DEFINITION | Human sapiens ABC transporter MOAT-C (MOAT-C) mRNA, complete cds. | | |
| ACCESSION | AF104942 | | |
| VERSION | AF104942.1 | GI:4140697 | |
| KEYWORDS | human. | | |
| ORGANISM | human. | | |
| JOURNAL | Medline | | |
| REFERENCE | 2 (bases 1 to 5838) | | |
| AUTHORS | Belinsky,M.G., Babin,L.J., Balsara,B.B., Testa,J.R. and Kruh,G.D. | | |
| TITLE | Characterization of MOAT-C and MOAT-D, new members of the MRP/CMOAT subfamily of transporter proteins | | |
| JOURNAL | J. Natl. Cancer Inst. 90 (22), 1735-1741 (1998) | | |
| FEATURES | 99043302 | | |
| SOURCE | 2 (bases 1 to 5838) | | |
| gene | Belinsky,M.G., Babin,L.J. and Kruh,G.D. | | |
| CDS | Direct Submission | | |
| | Submitted (04-NOV-1998) Medical Oncology, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA | | |
| | Location/Qualifiers | | |
| | 1..5838 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="3" | | |
| | /map="3q27" | | |
| | 1..5838 | | |
| | /gene="MOAT-C" | | |
| | 126..4439 | | |
| | /gene="MOAT-C" | | |
| | /note="MRP/CMOAT-related ABC transporter" | | |
| | /product="ABC transporter MOAT-C" | | |
| | /protein_id="AAD04169.1" | | |
| | /db_xref="GI:4140698" | | |
| | /translation="MKNDIDCKEIIISPGYRSYREHRSQTHRDREDSKFRFRRLP ECDDALEFAARAEGLSDASMSHSDRLIDEPHGKTHHGSAIKPIRTTSKHQPDV NAGLESCMTEFSWISLKLARVAHKELMEDWISLKHSSDYNCRLERLMOEELNE GPDASLRVWVICRRLILISYCLMIDTOLAGSGPAFWKHLELYQAESNIOYS LLIVAGILFEIYRSWSLALWALNTYRGVLRGAILTMARKILIKLNIEKSLGEI INISNGORHFEAAVAGSLAGGPVALIIGMIVNIIIGTGLGSAVFILFYPAAM FASLTIVFRKCYCAATDEROKNEVYLYIKFIKMYAPAFSOSYOK IREERRLI EKAGYFGCITGVAPYIVVIVASVTFESYHNLGFDLRAQAFTYVTFNSHTFALKYI PFSVKSISEASVAVDRKSLFLMEVHIKPKPASPHIKTEKNATLAWDSHSSIOA | | |

SPKLPKMKDKRKRASRGKKEKKEVROJORTHOAVLABEOKGHLLDSDERPSPEEBGKH
IHGHLRLQRTIHLSDLEIOBKLVLGICGSGKSTLSILIQMTLLEBSIASGT
FAYVAQOAMITLNLATLRLDNIETKEKEDERINSVLSCLPDLILPESDLIEGEB
ANLGGQORISLALYSDRSIYLIDPLSALAHVNHIFNSAIRHLKSKVLEFV
THOQYLVDCDEVIEMKEGCIETERGTHELNLUNGDAYTIFNNLLGCTPEVLENSK
ETSSOKKSODKPGTSVKKEKAVPEEQOLVOEESYGVYIOAAGP
LAFIVMALFPLANTGTAFTWMLSYMKGSGNTYTRGNETSGSOSDKMPHOY
ASIVALSMAVLLIKAIRGVYFVKTILABASSRLDELFRRILRSPKFEPTTPGRIL
NRESKDEVDVRLPFOEMFIQNVILVFCVGMAGVFWPLVAVGLVLFVSLH
VSRVILRELKRLDNIQSPFLSHITSIOGLATIHANXGOEFLHRQELLDNDQAF
FLFCAMKMLAVRLDLISALITTTGLMIVLMHQIPYAVGLIAYVQLTGJFOF
VRLASEEAFRTSEVERINHYIKTSLLEAPARIKNKAPSPDESEYFENAEKMYRE
NLPLVKKSFPTIKPREKIGIVRGSGSKSLGMAPFLVELSGCIRIDGVNRSDIG
LADRSKSLIIPQEPLEFSVRNSLNDPNSOYTEDQIMDALERTHMEKCIQOLPKLE
SEVENGDNEFSVGEROLICIAIRALRHKKIILIDBATRAMTEPDLIOETIRERPAD
CTMLTIAHRLHTVLJSDRIWLAQGVVEFDPVLSLNSDSRFTAFMAENKAVAK
G^a

BASE COUNT 1422 a 1432 c 1509 g 1475 t
ORIGIN

Query Match 98.9%; Score 4792.6; DB 88; Length 5838;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 4809; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1 GGCTATGCTGGAGCGGTGTTGAGCGGCTGGCGGTTGTCTGGAGAGGGGCGAG 60
DB 11 GGCTATGCTGGAGCGGTGTTGAGCGGCTGGCGGTTGTCTGGAGAGGGGCGAG 70
QY 61 GAATTCGATGTAAGTAACAGTCTGTGAGCCCTGGAACCTCACTCAGAAAGATGAA 120
DB 71 GAATTCGATGTAAGTAACAGTCTGTGAGCCCTGGAACCTCACTCAGAAAGATGAA 130
QY 121 GGATTCGATGTAAGTAACAGTCTGTGAGCCCTGGAACCTCACTCAGAAAGATGAA 180
DB 131 GGATTCGATGTAAGTAACAGTCTGTGAGCCCTGGAACCTCACTCAGAAAGATGAA 190
QY 181 GGAACACAGACACTTCTGGGAGCAGACAGACCGTGAAGATTCCAGTTGAGAGAACTCG 240
DB 191 GGAACACAGACACTTCTGGGAGCAGACAGACCGTGAAGATTCCAGTTGAGAGAACTCG 250
QY 241 ACCGTTGGAATCCCAAGATCCCTTGGAAACAGACGCCGAGGGGCTCTCTCTTGA 300
DB 251 ACCGTTGGAATCCCAAGATCCCTTGGAAACAGACGCCGAGGGGCTCTCTCTTGA 310
QY 301 TCCCTTCAGATGCTCTCAGCTCAGAAATCTGATGAGAGCATCCCAAGAGAAATGACA 360
DB 311 TCCCTTCAGATGCTCTCAGCTCAGAAATCTGATGAGAGCATCCCAAGAGAAATGACA 370
QY 361 TCAATGAGCTGAGTGTCTGAGCCATCCGACTACTTCAAAACACAGACCCAGTGA 420
DB 371 TCAATGAGCTGAGTGTCTGAGCCATCCGACTACTTCAAAACACAGACCCAGTGA 430
QY 421 CAATGCTGGGCTTTTTCCTGTATGACTTTTCGTGGCTTCTCTCTGCCCCGTGTGAC 480
DB 431 CAATGCTGGGCTTTTTCCTGTATGACTTTTCGTGGCTTCTCTCTGCCCCGTGTGAC 490
QY 481 CCACAGAAGAGGGGAGCTTCAATGAGAGAGCTGTGTCTGTCCAAACAGAGCTTTC 540
DB 491 CCACAGAAGAGGGGAGCTTCAATGAGAGAGCTGTGTCTGTCCAAACAGAGCTTTC 550
QY 541 TCAAGTGAAGCTCAGAGAGCTAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 551 TCAAGTGAAGCTCAGAGAGCTAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 610
QY 601 AGACGCTGCTTCCCTGCGAGAGGTTGTGATCTTCTCCACACAGAGCTATCTGTGC 660
DB 611 AGACGCTGCTTCCCTGCGAGAGGTTGTGATCTTCTCCACACAGAGCTATCTGTGC 670
QY 661 CATGCTGTGCTGATGATACGACAGCTGCTGCTTCACTGAGACAGCCTTCAATGATGAA 720
DB 671 CATGCTGTGCTGATGATACGACAGCTGCTGCTTCACTGAGACAGCCTTCAATGATGAA 730

QY 721 ACACCTTTGAGATATACCCAGAGCAAGAGCTTAACCTGACATGACAGCTTGTAGT 780
DB 731 ACACCTTTGAGATATACCCAGAGCAAGAGCTTAACCTGACATGACAGCTTGTAGT 790
QY 781 GCTGGGCTTCCCTCCAGAGGAAATGCTGGGCTTCTGTGCTTGTGACATGACTTGGCAT 840
DB 791 GCTGGGCTTCCCTCCAGAGGAAATGCTGGGCTTCTGTGCTTGTGACATGACTTGGCAT 850
QY 841 GAATTAACGAAACCGGTGCTCCGCTGCGGGGGCCATCTCAACATGAGCATTTAAGAGAT 900
DB 851 GAATTAACGAAACCGGTGCTCCGCTGCGGGGGCCATCTCAACATGAGCATTTAAGAGAT 910
QY 901 CCTTAAGTTAAAGAAATTAAGAGAAATCCCTGGGTGAGCTCATCAATTTGCTCAA 960
DB 911 CCTTAAGTTAAAGAAATTAAGAGAAATCCCTGGGTGAGCTCATCAATTTGCTCAA 970
QY 961 CGATGGGAGAGAAATGTTGAGCAGACAGCCGTTGGACAGCTGTGCTGAGAGACCGT 1020
DB 971 CGATGGGAGAGAAATGTTGAGCAGACAGCCGTTGGACAGCTGTGCTGAGAGACCGT 1030
QY 1021 TGTTCACATCTTAGCATGATTTATATGTAATTAATTCGGGAGCCACAGAGCTTCTGGG 1080
DB 1031 TGTTCACATCTTAGCATGATTTATATGTAATTAATTCGGGAGCCACAGAGCTTCTGGG 1090
QY 1081 ATCAGCTGTTTATATCTCTTTTACCCAGCAATGATGTTGCATCAGGCTCACAGCATA 1140
DB 1091 ATCAGCTGTTTATATCTCTTTTACCCAGCAATGATGTTGCATCAGGCTCACAGCATA 1150
QY 1141 TTTTCAGGAGAAATGCGTGGCCGCCACGAGATGACGTGTCCAGAAAGATGAAGTTCT 1200
DB 1151 TTTTCAGGAGAAATGCGTGGCCGCCACGAGATGACGTGTCCAGAAAGATGAAGTTCT 1210
QY 1201 TACTTACATTAATTTATCAAAATGATATGCTGGGTCAAAGCATTTTCCAGAGTTCA 1260
DB 1211 TACTTACATTAATTTATCAAAATGATATGCTGGGTCAAAGCATTTTCCAGAGTTCA 1270
QY 1261 GAAATATCCGAGAGAGAGAGCTCGGATATTGGAAGAAAGCGGGGTACTTCCAGAGCATC 1320
DB 1271 GAAATATCCGAGAGAGAGAGCTCGGATATTGGAAGAAAGCGGGGTACTTCCAGAGCATC 1330
QY 1321 TGTGGGTGTGCTCCATTTGTGTGTGTATGTCAGACGCTGTGACCTTCTGTTCATAT 1380
DB 1331 TGTGGGTGTGCTCCATTTGTGTGTGTATGTCAGACGCTGTGACCTTCTGTTCATAT 1390
QY 1381 GACCTTGGGCTTCAGATGACAGAGCAAGAGCTTTCAGAGGTGACAGCTTCAATTC 1440
DB 1391 GACCTTGGGCTTCAGATGACAGAGCAAGAGCTTTCAGAGGTGACAGCTTCAATTC 1450
QY 1441 CATGACTTTTCTTGAAGTAACACCGTTTCACTAAGTCCCTCTCAGAGCCCTCACT 1500
DB 1451 CATGACTTTTCTTGAAGTAACACCGTTTCACTAAGTCCCTCTCAGAGCCCTCACT 1510
QY 1501 GCGTGTGACAGATTTAAGAGTTTGTCTTAATGAGAGGTTTCACTAGATAAGAACAA 1560
DB 1511 GCGTGTGACAGATTTAAGAGTTTGTCTTAATGAGAGGTTTCACTAGATAAGAACAA 1570
QY 1561 ACCAGCAATCTCTCCATCAATCAATAGAGATGAGAAATGACACCTTGGCTGGAGCTCC 1620
DB 1571 ACCAGCAATCTCTCCATCAATCAATAGAGATGAGAAATGACACCTTGGCTGGAGCTCC 1630
QY 1621 CCACATCAATATCCAGAAATCTGCCCAAGCTGACCCCAAAATGAGAAAGAGAGAGG 1680
DB 1631 CCACATCAATATCCAGAAATCTGCCCAAGCTGACCCCAAAATGAGAAAGAGAGAGG 1690
QY 1681 TTCAGAGGCAAGAAAGAGAGAGTGAAGAGCTGACAGCCACTGAGCATCAGCGGTGCT 1740
DB 1691 TTCAGAGGCAAGAAAGAGAGAGTGAAGAGCTGACAGCCACTGAGCATCAGCGGTGCT 1750
QY 1741 GGCAGAGCAGAAAGGCACTCTCTCTGAGACAGTGAAGAGGCGCCAGTCCGAAAGAGA 1800
DB 1751 GGCAGAGCAGAAAGGCACTCTCTCTGAGACAGTGAAGAGGCGCCAGTCCGAAAGAGA 1810
QY 1801 AGAAGCAGACATATCCACTGGGCACTGCGCTTACAGAGAGACATCTCACAGCATGCA 1860

```
Db 1811 AGAAGGAGACATCCACCTGGGCGACCTGCTTACAGAGGACACTGCACACATCGA 1870
Qy 1861 TCTGAGATCCAGAGGGTAAACTGTTGGAATCTGGGGCGAGTGGGAAATGGAAAAAC 1920
Db 1871 TCTGAGATCCAGAGGGTAAACTGTTGGAATCTGGGGCGAGTGGGAAATGGAAAAAC 1930
Qy 1921 CTCTCATTTGACGCATTTTATGAGCCAGATGACGCTTCTAGAGGGGAGCATTTGCAATCG 1980
Db 1931 CTCTCATTTGACGCATTTTATGAGCCAGATGACGCTTCTAGAGGGGAGCATTTGCAATCG 1990
Qy 1981 TGAACCTTGCTTATATGTTGGCCGAGAGGCTGATCTCTCAATGCTACTCTGAAAGACAA 2040
Db 1991 TGAACCTTGCTTATATGTTGGCCGAGAGGCTGATCTCTCAATGCTACTCTGAAAGACAA 2050
Qy 2041 CATCTGTTTGGAGGAATTTGATGAAGAAAGATACAACTCTGCTGACACAGCTGCTG 2100
Db 2051 CATCTGTTTGGAGGAATTTGATGAAGAAAGATACAACTCTGCTGACACAGCTGCTG 2110
Qy 2101 CCGTAGGCTGACCTGAGCCATTTCTCCAGCAGCAGCCTGACGAGATTTGAGAGCGAGG 2160
Db 2111 CCGTAGGCTGACCTGAGCCATTTCTCCAGCAGCAGCCTGACGAGATTTGAGAGCGAGG 2170
Qy 2161 AGCCAACTGAGCGGTGGGCGCCGAGAGATGAGCCTTGGCGGGCTGTATAGTGA 2220
Db 2171 AGCCAACTGAGCGGTGGGCGCCGAGAGATGAGCCTTGGCGGGCTGTATAGTGA 2230
Qy 2221 CAGAGATCTACATCTCTGAGCAGCAGCCTGAGCTTAAATGATCCCATGCGGCAACCA 2280
Db 2231 CAGAGATCTACATCTCTGAGCAGCAGCCTGAGCTTAAATGATCCCATGCGGCAACCA 2290
Qy 2281 CATCTTCAATAGTCTTATCCGGAAGCATCTCACTCAAGACAGTCTGTTTGTACCA 2340
Db 2291 CATCTTCAATAGTCTTATCCGGAAGCATCTCACTCAAGACAGTCTGTTTGTACCA 2350
Qy 2341 CCATTTACAGTACCTGCTGCTGATGATGATGATCTCTCAAGAGGCGCTATATAC 2400
Db 2351 CCATTTACAGTACCTGCTGCTGATGATGATGATCTCTCAAGAGGCGCTATATAC 2410
Qy 2401 GGAAGAGGAGCCCATGAGAGAACTGATGATTTAAATGATGATGATGATGATGATGAT 2460
Db 2411 GGAAGAGGAGCCCATGAGAGAACTGATGATTTAAATGATGATGATGATGATGATGAT 2470
Qy 2461 TAACCTGTGCTGGGAGAGACAGCCGAGTTGAGATCAATTTCAAAAAAGAAACAGTGG 2520
Db 2471 TAACCTGTGCTGGGAGAGACAGCCGAGTTGAGATCAATTTCAAAAAAGAAACAGTGG 2530
Qy 2521 TTTCAGAGAGATGATCAAGCAAGGGGCTTAAACAGAGATCAATTAAGAGAGAAAGGC 2580
Db 2531 TTTCAGAGAGATGATCAAGCAAGGGGCTTAAACAGAGATCAATTAAGAGAGAAAGGC 2590
Qy 2581 AGTAAAGCCAGAGAGAGGAGCTTGTGACGTGGAAGAGAAAGGAGGAGTTGACAGTGC 2640
Db 2591 AGTAAAGCCAGAGAGAGGAGCTTGTGACGTGGAAGAGAAAGGAGGAGTTGACAGTGC 2650
Qy 2641 CTGCTGATGATATATGCTGCTCAATCCAGGCTGCTGGGGGCGCCCTTGGCATCTGCTGAT 2700
Db 2651 CTGCTGATGATATATGCTGCTCAATCCAGGCTGCTGGGGGCGCCCTTGGCATCTGCTGAT 2710
Qy 2701 TATGAGCCCTTTTCAATGCTGATGATGAGGAGCAGCCCTTCAAGCACTGCTGCTGATGAT 2760
Db 2711 TATGAGCCCTTTTCAATGCTGATGATGAGGAGCAGCCCTTCAAGCACTGCTGCTGATGAT 2770
Qy 2761 CTGATCAAGCAAGAGAGGAGGAGCAGCAGTGTGATCTGAGGAGAGAGAGCAGCTGCTGAG 2820
Db 2771 CTGATCAAGCAAGAGAGGAGGAGCAGCAGTGTGATCTGAGGAGAGAGAGCAGCTGCTGAG 2830
Qy 2821 TGAACAGATGAAGAGCAATCTCTCATATGCACTACTATGCGAGCATCTACGCGCTCTCAT 2880
Db 2831 TGAACAGATGAAGAGCAATCTCTCATATGCACTACTATGCGAGCATCTACGCGCTCTCAT 2890
Qy 2881 GGCAGTATGCTGATCTGAAAGGCAATTCGAGGAGTGTCTTTGCAAGGGCAGAGCTGCG 2940
Db 2891 GGCAGTATGCTGATCTGAAAGGCAATTCGAGGAGTGTCTTTGCAAGGGCAGAGCTGCG 2950
Qy 2941 AGCTTCTCCCGGCTGATGACAGAGCTTTTCCGAAGATCTTCCGAAGCCCTATGAAGTT 3000
Db 2951 AGCTTCTCCCGGCTGATGACAGAGCTTTTCCGAAGATCTTCCGAAGCCCTATGAAGTT 3010
Qy 3001 TTTTGCACGAGCCCGACAGGAGGAGATTCTCAAGATTTTCCAAAGCATGATGAAGT 3060
Db 3011 TTTTGCACGAGCCCGACAGGAGGAGATTCTCAAGATTTTCCAAAGCATGATGAAGT 3070
Qy 3061 TGAGGTGCGGCTGCGGTTCCAGGCGGAGATGTTATCATCAGAACGTTATCCGCTGCTT 3120
Db 3071 TGAGGTGCGGCTGCGGTTCCAGGCGGAGATGTTATCATCAGAACGTTATCCGCTGCTT 3130
Qy 3121 CTGTGTGGAATGATTCGAGAGATCTTCCGAGTTCCTGTTGCGAGTGGGGCCCTTGT 3180
Db 3131 CTGTGTGGAATGATTCGAGAGATCTTCCGAGTTCCTGTTGCGAGTGGGGCCCTTGT 3190
Qy 3181 CATCTCTTTTCACTCTGACATTTGCTCCAGGCTCTGATTTGGGAGCTGAAAGCTCT 3240
Db 3191 CATCTCTTTTCACTCTGACATTTGCTCCAGGCTCTGATTTGGGAGCTGAAAGCTCT 3250
Qy 3241 GGACAAATTCACGAGTACACCTTCTCTCCACATCAGTCCAGATACAGGCGCTTGC 3300
Db 3251 GGACAAATTCACGAGTACACCTTCTCTCCACATCAGTCCAGATACAGGCGCTTGC 3310
Qy 3301 CACCATCCAGCCTACATAAAGGAGAGATTCTGACAGATACGAGAGCTGCTGGA 3360
Db 3311 CACCATCCAGCCTACATAAAGGAGAGATTCTGACAGATACGAGAGCTGCTGGA 3370
Qy 3361 TGACAAACAGCTCTTTTGTGTTTGTAGCTGACATGGGGGCTGCTGCTGCTGCTGCT 3420
Db 3371 TGACAAACAGCTCTTTTGTGTTTGTAGCTGACATGGGGGCTGCTGCTGCTGCTGCT 3430
Qy 3421 GGACCTATCAGATTCGCGCTCATCACACACAGGCGCTGATGATCTTATGACAGG 3480
Db 3431 GGACCTATCAGATTCGCGCTCATCACACACAGGCGCTGATGATCTTATGACAGG 3490
Qy 3481 GCAGATTCGCCAGCCTATGCGGGCTCTGCGCATCTTATGCTCCAGTTAACGGGCT 3540
Db 3491 GCAGATTCGCCAGCCTATGCGGGCTCTGCGCATCTTATGCTCCAGTTAACGGGCT 3550
Qy 3541 GTTCCAGTTTACGGTACAGTGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Db 3551 GTTCCAGTTTACGGTACAGTGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3610
Qy 3601 GATCAATCACTACATTTAAGACTGTCTTGAAGCACTGCCAGATTTAAGAAAGAGC 3660
Db 3611 GATCAATCACTACATTTAAGACTGTCTTGAAGCACTGCCAGATTTAAGAAAGAGC 3670
Qy 3661 TCCCTCCCTGACTGGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Db 3671 TCCCTCCCTGACTGGGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3730
Qy 3721 CCGAGAAACCTCCCTGCTGCTTAAAGAAATGATCTTCAGATCAAACTTAAGAGAA 3780
Db 3731 CCGAGAAACCTCCCTGCTGCTTAAAGAAATGATCTTCAGATCAAACTTAAGAGAA 3790
Qy 3781 GATTTGAGATGTTGGGGGAGAGAGATCAGGAGAGTCCGCTGGGAGTGGCCCTTCCG 3840
Db 3791 GATTTGAGATGTTGGGGGAGAGAGATCAGGAGAGTCCGCTGGGAGTGGCCCTTCCG 3850
Qy 3841 TCTGTGTGAGATTATCTGAGAGCTGCATCAGATGATGAGATGAGATGAGATGAG 3900
Db 3851 TCTGTGTGAGATTATCTGAGAGCTGCATCAGATGATGAGATGAGATGAGATGAG 3910
Qy 3901 CTTTGCAGACCTCCGAGCAAACTCTATCATCTCTCAAGAGCGGGGCTGCTCACTG 3960
Db 3911 CTTTGCAGACCTCCGAGCAAACTCTATCATCTCTCAAGAGCGGGGCTGCTCACTG 3970
Qy 3961 CACTGTGAGATCAATTTGAGAGCCCTTCAACAGATGACTGAAGAGAGATTTGGAGTGC 4020
Db 3971 CACTGTGAGATCAATTTGAGAGCCCTTCAACAGATGACTGAAGAGAGATTTGGAGTGC 4030
```

| | | | |
|----|------|---|------|
| QY | 244 | GTGGAAATGCCAAGATGCTTTGGAAACAGCAGCCGACGGGCGCTCTCTCTGATGTC | 303 |
| Db | 325 | GTTCGAATGCCAAGATGCTTTGGAAACAGCAGCCGACGGGCGCTCTCTGATGTC | 384 |
| QY | 304 | CTCCATGCATTCGAGCTCCAGGAATCCGGAATGAGGAGCAATCCCAAGGGAATACCATCA | 363 |
| Db | 385 | CTCCATGCATTCGAGCTCCAGGAATCCGGAATGAGGAGCAATCCCAAGGGAATACCATCA | 444 |
| QY | 364 | TGGCTTGAAGTCTCTGAAAGCCATCCGGAATACCTTCCAAACACAGCACCAGTGGACAA | 423 |
| Db | 445 | TGGCTTGAAGTCTCTGAAAGCCATCCGGAATACCTTCCAAACACAGCACCAGTGGACAA | 504 |
| QY | 424 | TGCTGGGCTTTTTCGTGATGACTTTTCGTGGCTTTCTCTCTGGCCCGTGTGGCCCA | 483 |
| Db | 505 | TGCTGGGCTTTTTCGTGATGACTTTTTCGTGGCTTTCTCTCTGGCCCGTGTGGCCCA | 564 |
| QY | 484 | CAGAAGAGGAGAGCTCTCAATGGAAAGCGTGTGTCTCTGTCCAGCAGCACTTTCTGCA | 543 |
| Db | 565 | CAGAAGAGGAGAGCTCTCAATGGAAAGCGTGTGTCTCTGTCCAGCAGCACTTTCTGCA | 624 |
| QY | 544 | CGTGAACGCGAAGAGACTAGAGAGACTGTGGCAAGAAAGCTGAATGAATTTGGGCGACA | 603 |
| Db | 625 | CGTGAACGCGAAGAGACTAGAGAGACTGTGGCAAGAAAGCTGAATGAATTTGGGCGACA | 684 |
| QY | 604 | CGCTGCTTCCCTGGCAAGGGTGTGTGATCTTCTGCGCAGCAGGAGCTATCCTGTGCAT | 663 |
| Db | 685 | CGCTGCTTCCCTGGCAAGGGTGTGTGATCTTCTGCGCAGCAGGAGCTATCCTGTGCAT | 744 |
| QY | 664 | CGTGTGCTGTATGATCAACGCAAGCTGGCTTGCTCAGTGAAGCAACGCTTCATGTGAACA | 723 |
| Db | 745 | CGTGTGCTGTATGATCAACGCAAGCTGGCTTGCTCAGTGAAGCAACGCTTCATGTGAACA | 804 |
| QY | 724 | CCTCTGGAGATATCCCAAGCAACAGAGCTAATACCTGCAGCTACAGCTTTCTGTATGCT | 783 |
| Db | 805 | CCTCTGGAGATATCCCAAGCAACAGAGCTAATACCTGCAGCTAATGCTTTCTGTATGCT | 864 |
| QY | 784 | GGGGCTCCTCTGACGGAAATCGTGCAGCTTGTGTGCTGTGCACTGACTTGGGCAATTGAA | 843 |
| Db | 865 | GGGGCTCCTCTGACGGAAATCGTGCAGCTTGTGTGCTGTGCACTGACTTGGGCGTTGAA | 924 |
| QY | 844 | TTTACCGAACCGGCTCCGCTTGGCGGGGGCCATCTTACCATGGCAATTTTAAAGATGCT | 903 |
| Db | 925 | TTTACCGAACCGGCTCCGCTTGGCGGGGGCCATCTTACCATGGCAATTTTAAAGATGCT | 984 |
| QY | 904 | TTAAGTTAAAGAACTTTAAAGAAATCCCGGGGTGAGCTCAATCAATTTGCTCCAGCA | 963 |
| Db | 985 | TTAAGTTAAAGAACTTTAAAGAAATCCCGGGGTGAGCTCAATCAATTTGCTCCAGCA | 1044 |
| QY | 964 | TGGGCAAGAAATGTTTGAAGGAGAGCGCTTGGCAGAGCTCTGGCTGGAGAGACCGTGT | 1022 |
| Db | 1045 | TGGGCAAGAAATGTTTGAAGGAGAGCGCTTGGCAGAGCTCTGGCTGGAGAGACCGTGT | 1104 |
| QY | 1024 | TGCCATCTTAAAGCATGATTTATTAATGTAATTTCTGGGACCAACAGGCTCTCGGGATC | 1083 |
| Db | 1105 | TGCCATCTTAAAGCATGATTTATTAATGTAATTTCTGGGACCAACAGGCTCTCGGGATC | 1164 |
| QY | 1084 | AGCTGTTTTTATTCCTCTTTTAAACCAGAAATGATGTTTGAATACAGGCTCTACAGCATATTT | 1144 |
| Db | 1165 | AGCTGTTTTTATTCCTCTTTTAAACCAGAAATGATGTTTGAATACAGGCTCTACAGCATATTT | 1224 |
| QY | 1144 | CAGGAGAAATGCGTGGCGCGCAAGGATGAAGCTGTCCACAAATGAATGAATGAATGTTTAC | 1203 |
| Db | 1225 | CAGGAGAAATGCGTGGCGCGCAAGGATGAAGCTGTCCACAAATGAATGAATGAATGTTTAC | 1284 |
| QY | 1204 | TTTACATTAATTTATCABAATGTATGCTTGGCTGCAAAACATTTTCTCAGAGTGTTCAGAA | 1263 |
| Db | 1285 | TTTACATTAATTTATCABAATGTATGCTTGGCTGCAAAACATTTTCTCAGAGTGTTCAGAA | 1344 |
| QY | 1264 | AATCCGCAAGAGAGAGGTGAGATTTGAAAAAGCGGGTACTTCCAGACAGATCAAGCT | 1322 |
| Db | 1345 | AATCCGCAAGAGAGAGGTGAGATTTGAAAAAGCGGGTACTTCCAGACAGATCAAGCT | 1404 |
| QY | 1324 | GGGTGTGGCTCCCATGTGGTGGATGTGCACGCTGTGATACCTTCTCTGTCATATGATC | 1383 |

| | | | |
|----|------|---|------|
| Db | 1405 | GGGtGTGgCTCCCAATGTGTGGTGTGATGTGCAGACGGTGTGACCTTCTCTGTTCAATATAC | 1464 |
| Qy | 1384 | CCGTGGCGTTCGATCTGTGCACACACACAGGCTTTCACAGTGTGTGACAGTCTTCAATTCCAT | 1443 |
| Db | 1465 | CTTGGGCTTGCAATGTGCACACACACAGGCTTTCACAGTGTGTGACAGTCTTCAATTCCAT | 1524 |
| Qy | 1444 | GACTTTTGCTTTGAAAGTAACACCGTTTTCAGTAAAGTCCCTCAGAAAGCTCAGTGGC | 1503 |
| Db | 1525 | GACTTTTGCTTTGAAAGTAACACCGTTTTCAGTAAAGTCCCTCAGAAAGCTCAGTGGC | 1584 |
| Qy | 1504 | TGTTTGCACAGATTTAAGAGTTTGTTCATATGCAAGAGTTTCACATGATTAAGAACAAAC | 1563 |
| Db | 1585 | TGTTTGCACAGATTTAAGAGTTTGTTCATATGCAAGAGTTTCACATGATTAAGAACAAAC | 1644 |
| Qy | 1564 | AGCCAGTCCACATCAGATTAAGATTAAGAAATATGCACTTGGCATGGGATCCTCCCA | 1623 |
| Db | 1645 | AGCCAGTCCACATCAGATTAAGATTAAGAAATATGCACTTGGCATGGGATCCTCCCA | 1704 |
| Qy | 1624 | CTCCAGATCCAGAAATCGGCCCAAGCTGACCCCAAAATAGAAAAAGAACAGAGGGCTTC | 1683 |
| Db | 1705 | CTCCAGATCCAGAAATCGGCCCAAGCTGACCCCAAAATAGAAAAAGAACAGAGGGCTTC | 1764 |
| Qy | 1684 | CAGGGGCAAGAAAGAGAAGGTGAGGACGCTGCAGCGCATGAGCATCAGCGGTGCTGGC | 1743 |
| Db | 1765 | CAGGGGCAAGAAAGAGAAGGTGAGGACGCTGCAGCGCATGAGCATCAGCGGTGCTGGC | 1824 |
| Qy | 1744 | AGAGCAGAAAGGCCACTCTCTCTGCAAGTACACAGCGGCCCATCTCCGAAAGAGGAAGA | 1803 |
| Db | 1825 | AGAGCAGAAAGGCCACTCTCTCTGCAAGTACACAGCGGCCCATCTCCGAAAGAGGAAGA | 1884 |
| Qy | 1804 | AGGCAAGACATCCACTGGGGCCACGCGCTTACAGAGAACACTGCACAGCATGATCT | 1863 |
| Db | 1885 | AGGCAAGACATCCACTGGGGCCACGCGCTTACAGAGAACACTGCACAGCATGATCT | 1944 |
| Qy | 1864 | GGAGATCCAAAGAGGTAACCTGTTGGAACTCTGCGCAGTGTGGGAATGTGAAAAAACCCTC | 1923 |
| Db | 1945 | GGAGATCCAAAGAGGTAACCTGTTGGAACTCTGCGCAGTGTGGGAATGTGAAAAAACCCTC | 2004 |
| Qy | 1924 | TCTCATTTCAAGCTTTTAAAGCCAGATGACGCTTCTAAGAGGCGACGATTTGCAATACATGG | 1983 |
| Db | 2005 | TCTCATTTCAAGCTTTTAAAGCCAGATGACGCTTCTAAGAGGCGACGATTTGCAATACATGG | 2064 |
| Qy | 1984 | AACCTTGCTTATGTGGCCAGCAGGCGCTGGATCTCCAATGCTACTCTAGAGACAAAT | 2043 |
| Db | 2065 | AACCTTGCTTATGTGGCCAGCAGGCGCTGGATCTCCAATGCTACTCTAGAGACAAAT | 2124 |
| Qy | 2044 | CCGTGTTGGGAAGAAATATGATGAAGAAATCAACTCTGTGCTGAACAGCTGCTGCT | 2103 |
| Db | 2125 | CCGTGTTGGGAAGAAATATGATGAAGAAATCAACTCTGTGCTGAACAGCTGCTGCT | 2184 |
| Qy | 2104 | GAGGCTGACCTGGCCATTTCTCCACAGCCACTGCAGAGATTTGAGAGCGAGAGC | 2163 |
| Db | 2185 | GAGGCTGACCTGGCCATTTCTCCACAGCCACTGCAGAGATTTGAGAGCGAGAGC | 2244 |
| Qy | 2164 | CAACCTGAGGCTGGGAGAGGCCAGAGGATCACCTTGGCCCGGCTTATATAGGACAG | 2223 |
| Db | 2245 | CAACCTGAGGCTGGGAGAGGCCAGAGGATCACCTTGGCCCGGCTTATATAGGACAG | 2304 |
| Qy | 2224 | GAGCATACATCCCTGGACACACCCCTCATGTGCTTATAGTGCATGTGGGCAACCAT | 2283 |
| Db | 2305 | GAGCATACATCCCTGGACACACCCCTCATGTGCTTATAGTGCATGTGGGCAACCAT | 2364 |
| Qy | 2284 | CTTCAATATGTCTATCCGGAAACATCTCAAGTCCAAAGACAGTTCTGTGTTATACCCACA | 2343 |
| Db | 2365 | CTTCAATATGTCTATCCGGAAACATCTCAAGTCCAAAGACAGTTCTGTGTTATACCCACA | 2424 |
| Qy | 2344 | GTTACAGACCTGTGTACAGTGTGTAAGTGAATCTTCATGAAGAGAGGCTGTATTAACGA | 2403 |
| Db | 2425 | GTTACAGACCTGTGTACAGTGTGTAAGTGAATCTTCATGAAGAGAGGCTGTATTAACGA | 2484 |
| Qy | 2404 | AAGAGGACCCCATAGAGAACTGATGAATTTAAATGTGACTATGCTACCATTTTAAATPA | 2463 |

QY 4624 TGTAAACAAATTTAGTTTGTCTTAAATTCACCTCTAAAGGCTCAGGACCGTTAT 4683
|||||
DB 4704 TGTAAACAAATTTAGTTTGTCTTAAATTCACCTCTAAAGGCTCAGGACCGTTAT 4763
QY 4684 TATTAATGTATCAGAGCCCTTAATGAAGCTTTATACCTAGCTATATATATAT 4743
|||||
DB 4764 TATTAATGTATCAGAGCCCTTAATGAAGCTTTATACCTAGCTATATATATAT 4823
QY 4744 TCTGTACATACCTATATTTTACATGAAATGTAAAGCTTTATTTATATTAATAG 4803
|||||
DB 4824 TCTGTACATACCTATATTTTACATGAAATGTAAAGCTTTATTTATATTAATAG 4883
QY 4804 CACTGTGCTAAAAAAA 4819
|||||
DB 4884 CACTGTGCTATATACA 4899
RESULT 4
HS083661 5826 bp mRNA PRI 21-JUN-2000
LOCUS Homo sapiens multidrug resistance protein 5 (MRP5) mRNA, complete
DEFINITION cds.
ACCESSION U83661
VERSION U83661.2 GI:5685863
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 3765 to 5822)
AUTHORS Kool,M., de Haas,M., Scheffer,G.L., Scheper,R.J., van Eijk,M.J.,
Juijn,J.A., Baas,F. and Borst,P.
TITLE Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5,
homologues of the multidrug resistance-associated protein gene
(MRP1), in human cancer cell lines
JOURNAL Cancer Res. 57 (16), 3537-3547 (1997)
MEDLINE 97413640
REFERENCE 2 (bases 1 to 5826)
AUTHORS Wijnholds,J., Mol,C.A., van Deemter,L., de Haas,M., Scheffer,G.L.,
Baas,F., Beijnen,J.H., Scheper,R.J., Hulse,S., De Clercq,E.,
Balzarini,J. and Borst,P.
TITLE Multidrug-resistance protein 5 is a multispecific organic anion
transporter able to transport nucleotide analogs
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (13), 7476-7481 (2000)
MEDLINE 20319051
REFERENCE 3 (bases 3765 to 5822)
AUTHORS Kool,M., Wijnholds,J., de Haas,M., Mol,C.A.A.M., Ponne,N.J.,
Baas,F. and Borst,P.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-1997) Molecular Biology, Netherlands Cancer
Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands
REFERENCE 4 (bases 1 to 5826)
AUTHORS Wijnholds,J., Mol,C.A.A.M., Ponne,N.J., Baas,F. and Borst,P.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Molecular Biology, Netherlands Cancer
Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands
REMARK Sequence update by submitter
COMMENT On Aug 3, 1999 this sequence version replaced gi:2439973.
FEATURES
source
1..5826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="7.69 cR from WI-6365"
/tissue_type="brain"
/dev_stage="fetus"
1..5826
/gene="MRP5"
120..4433
/gene="MRP5"
/note="ABC transporter protein; MOAT-C; PABCI1"
/codon_start=1
/product="multidrug resistance protein 5"

/protein_id="AA071758.2"
/db_xref="GI:5685864"
/translation="MKDIDIGKEVYIIIPSGYRSVRERTSTSGTHRDREDSKERRRPL
ECDDALEETARABGLSDLSMSHQILIDDEHKKGYHRLSLAKIRITTSKQHYD
NAGLFSCMTFSWLISLARVAHKRGELSMEDVSLSHESDVNKLRELTQATELQSL
GPDASLRVWMIJECRRLILSLVCLMTQLDAPSGAPFVWKLRELTQATELQSL
LLVLVLGLLEIVRSMSLALTMALNRTGRLGALILTMFKILILKILNKESLDEL
INICSDNGORMPFAAVGSLLAGPVVAILGMIYNYILIGPTGFLSAVILFYPM
FASRLTVAFFRRKCVANRDERVKMNEVLYIKIKIYAAWKAQSOSQKIFREERL
EKAGYQSITVGVAPIVVYIASVYTSVNHITLFKPDITAAQAPVYVYFNSMTALVT
PFSVKSLSEASVAVDRKSLFLMEVHMINKRASPITKEMNATLWDSSTION
SPRLTPKMKDKRASRKKERKQRLQRTHEQAVLAQKGLLDSDERSPEDEKX
IHGLRLRLTSLIDLEIDEGKLVGSGVSGSKSLISAILGOMTLLENKSLIGT
FAYVADQAWILNATLNDNLIFGKEYDERNSVNSCCLPDAIPLSPDLTIGERG
ANLSGGOORILSARALYSRSLYIIDDPSALDAHGNHIFNSAIRKHLSKTVLF
THLOQTLVDDDEYIPAKGEGITRGRHLEELNNGXATTFNNLLGEPFVINSK
ETISGSKSDODKPTGTSVKEKAVRKEBGLVOLERKGGSPVSWYGYTDAAGP
LAPLVIALFPLMNVGSAFSTWMLSWIKGSGNTVYTRNENVSQSMKDNHMOY
ASIVALSMAVMLIKAIRGVVFKGTLRASSRLHDELFRILRSPMKFPDTPPTRL
NRESKMDDEVADLPLQAEFIONVLIVFCVGMIGVFPWFLVAVGLIFLSVLI
VSRVILIRELRLDNLITQSPLSHTSSIOGLATIHAVNKGOELHRYOELDNOAF
FLTPCARMVLAVRUDLISLITTTGMLTYMHGQIPRAYAGLISVAOVLGLOFT
VRUSETEARFTVERINHAIKTLSEAPRIINKRASPMPDQGEVTFENAKRIE
NLPLVLKKSFTLKPKEKIGIVGTGSGKSLGMALFRLVELGGCKRIDGVRLSDIG
LADLRSLIITPOEPVLFSGTVRSNLDPEQYEDQIMDALERTHEKICIAQPLKLE
SEVMDNDFNSVERDLICARALLRICKILIIDEAATAADTETDLILOETIREAFD
CTMLTIAHRLHYVIGSDRIWVLAQGOVYEDPFSVLSNDSRFYAMFAAENKVAVK
G"

BASE COUNT 1404 a 1427 c 1511 g 1474 t 10 others
ORIGIN

Query Match 98.8%; Score 4788; DB 97; Length 5826;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4802; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 12 GGGAGCGGTGTGAGCGCGCTGCCGCGTTCCTGAGACAGGCGCAGAAATCTGATG 71
|||||
DB 16 GGGAGCGGTGTGAGCGCGCTGCCGCGTTCCTGAGACAGGCGCAGAAATCTGATG 75
QY 72 TGAACCTAACAGTCTGTGAGCCCTGGAACCTCCACTCAGAGAAATGAGATATCGACA 131
|||||
DB 76 TGAACCTAACAGTCTGTGAGCCCTGGAACCTCCACTCAGAGAAATGAGATATCGACA 135
QY 132 TAGGAAAAGATATATATATATCCCACTCCTGGGTATAGAAAGTGAAGAGAGAACCGACA 191
|||||
DB 136 TAGGAAAAGAT 195
QY 192 CTTCTGGAGCCACAGAGACCGTGAAGATTCAGATTCAGAGAACTGACCGTTGGAAAT 251
|||||
DB 196 CTTCTGGAGCCACAGAGACCGTGAAGATTCAGATTCAGAGAACTGACCGTTGGAAAT 255
QY 252 GCCAAGATGCTTGGAAAACAGACGCCGAGGCGCTCTCTCTTGATGCTTCATATG 311
|||||
DB 256 GCCAAGATGCTTGGAAAACAGACGCCGAGGCGCTCTCTCTTGATGCTTCATATG 315
QY 312 ATTCTCAGCTAGAAATCTGATGAGAGCATCCCAAGGAGAAATGATCATATGCTTGA 371
|||||
DB 316 ATTCTCAGCTAGAAATCTGATGAGAGCATCCCAAGGAGAAATGATCATATGCTTGA 375
QY 372 GTGCTGTGAAGCCCATCGGACTACTTGTCAAAACACAGACCCAGTGGACAAATCTGAGC 431
|||||
DB 376 GTGCTGTGAAGCCCATCGGACTACTTGTCAAAACACAGACCCAGTGGACAAATCTGAGC 435
QY 432 TTTTCTCTGTATGACTTTTCTGTGGCTTCTCTGTGGCCGTGTGGCCACAGAAAG 491
|||||
DB 436 TTTTCTCTGTATGACTTTTCTGTGGCTTCTCTGTGGCCGTGTGGCCACAGAAAG 495
QY 492 GGGAGCTCTCAATGAGAGAGCTGTGCTCTGTCACAGACAGATCTTTCAGAGTGAAT 551
|||||
DB 496 GGGAGCTCTCAATGAGAGAGCTGTGCTCTGTCACAGACAGATCTTTCAGAGTGAAT 555
QY 552 GCAGAGACTGTAGAGAGACTGTGCGAAGAGAGTGAATGAGTTGGGCGACAGCGCTGT 611

Db 2485 AAGAGCACCCTGAGAGACATGATTAATTTAAATGCTATGCTACCATTTTAAATA 2544
QY 2464 CCTGTTGCTGGAGAGACACCGCAGTTGAGATCAATTCAAAAAGAAACCAAGTGGTTC 2523
Db 2545 CCTGTTGCTGGAGAGACACCGCAGTTGAGATCAATTCAAAAAGAAACCAAGTGGTTC 2604
QY 2534 ACAGAGAAGTCAACAAGACAGAGGTCCTTAAACAGAGATCAATTAAGAGAAAGAGCACT 2583
Db 2605 ACAGAGAAGTCAACAAGACAGAGGTCCTTAAACAGAGATCAATTAAGAGAAAGAGCACT 2664
QY 2584 AAAGCAGAGAGAGAGGAGCTGTGACGTGAGAGAGAAAGGAGGAGGTCAGTGGCCCG 2643
Db 2665 AAAGCAGAGAGAGAGGAGCTGTGACGTGAGAGAGAAAGGAGGAGGTCAGTGGCCCG 2724
QY 2644 GTCACTATATGCTGCTCATATCCAGGCTGCTGGGGCCCTTGGCATTCCTGTTATAT 2703
Db 2725 GTCACTATATGCTGCTCATATCCAGGCTGCTGGGGCCCTTGGCATTCCTGTTATAT 2784
QY 2704 GGCCCTTTTCATGCTGAATGTAGGAGACACCGCTTCAGACACCTGGTGGTGAATTACTG 2763
Db 2785 GGCCCTTTTCATGCTGAATGTAGGAGACACCGCTTCAGACACCTGGTGGTGAATTACTG 2844
QY 2764 GATCAAGAGAGAGAGAGAGAGACACCACTGTGACTGAGAGAGAGAGAGAGAGAGAGTA 2823
Db 2845 GATCAAGTA 2904
QY 2824 CAGCATGAAGTA 2883
Db 2905 CAGCATGAAGTA 2964
QY 2884 AGTCATGCTGATCTGAAAG 2943
Db 2965 AGTCATGCTGATCTGAAAG 3024
QY 2944 TTTCCTCCCGGCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3003
Db 3025 TTTCCTCCCGGCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3084
QY 3004 TGACAGTA 3063
Db 3085 TGACAGTA 3144
QY 3064 CGTGGGCTGCGGCTTCAG 3123
Db 3145 CGTGGGCTGCGGCTTCAG 3204
QY 3124 TTGGGAGTA 3183
Db 3205 TTGGGAGTA 3264
QY 3184 CCTTTTTCAGTCTGACATATGCTCCAGAGTCCGATTCGGAGAGTCAAGAGTCTGCA 3243
Db 3265 CCTTTTTCAGTCTGACATATGCTCCAGAGTCCGATTCGGAGAGTCAAGAGTCTGCA 3324
QY 3244 CAATATCAAGTA 3303
Db 3325 CAATATCAAGTA 3384
QY 3304 CATTCAGTA 3363
Db 3385 CATTCAGTA 3444
QY 3364 CAACCAAGTA 3423
Db 3445 CAACCAAGTA 3504
QY 3424 CCTCATCAGTA 3483
Db 3505 CCTCATCAGTA 3564
QY 3484 GATTCAGTA 3543
Db 3565 GATTCAGTA 3624

QY 3544 CCAAGTTACGCTGACAGTGGCATCTGAGACAGAGAGCTGATTCACCTGGTGAGAGAT 3603
Db 3625 CCAAGTTACGCTGACAGTGGCATCTGAGACAGAGAGCTGATTCACCTGGTGAGAGAT 3684
QY 3604 CAATCACTACATTAAGTA 3663
Db 3685 CAATCACTACATTAAGTA 3744
QY 3664 CTCCCTGAGTGGCCCGAGTA 3723
Db 3745 CTCCCTGAGTGGCCCGAGTA 3804
QY 3724 AGAAAACTCCCTCTGCTCTTAAAGAGATATCTTCACAGATCAAACTTAAAGAGAGAT 3783
Db 3805 AGAAAACTCCCTCTGCTCTTAAAGAGATATCTTCACAGATCAAACTTAAAGAGAGAT 3864
QY 3784 TGGCATTTGGGGCGGAGTA 3843
Db 3865 TGGCATTTGGGGCGGAGTA 3924
QY 3844 GGTGAGATATCTGAGAGGCTGATCAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGTA 3903
Db 3925 GGTGAGATATCTGAGAGGCTGATCAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGTA 3984
QY 3904 TGCCGACCTCCGAGTA 3963
Db 3985 TGCCGACCTCCGAGTA 4044
QY 3964 TGTGAGATTAATTTGGAGTA 4023
Db 4045 TGTGAGATTAATTTGGAGTA 4104
QY 4024 GGAGTA 4083
Db 4105 GGAGTA 4164
QY 4084 GATGAGTA 4143
Db 4165 GATGAGTA 4224
QY 4144 CCTGCTCCGCTGCTGATGAGATTCATTTAGATGAAGAGAGAGAGAGAGAGAGAGAGTA 4203
Db 4225 CCTGCTCCGCTGCTGATGAGATTCATTTAGATGAAGAGAGAGAGAGAGAGAGAGTA 4284
QY 4204 GACAGAGATTAATTTGAGTA 4263
Db 4285 GACAGAGATTAATTTGAGTA 4344
QY 4264 CATTCGCCATTCGCTGACAGAGGTTCTAGGCTCCGATGAGATTAAGAGAGAGAGTA 4323
Db 4345 CATTCGCCATTCGCTGACAGAGGTTCTAGGCTCCGATGAGATTAAGAGAGAGAGTA 4404
QY 4324 ACAGAGTGGTGGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTA 4383
Db 4405 ACAGAGTGGTGGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTA 4464
QY 4384 TGCCATGTTGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTA 4443
Db 4465 TGCCATGTTGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTA 4524
QY 4444 CGAAGTCTCTTTCTTTAGAGATTCGATTCCTGCTGGGGGGGGCCCTTCATCGCG 4503
Db 4525 CGAAGTCTCTTTCTTTAGAGATTCGATTCCTGCTGGGGGGGGCCCTTCATCGCG 4583
QY 4504 TCTCTCTACCGAAAGCTTGCCTTTCTGAGATTTATCTTTGCGAGAGAGATTCGAGATTCG 4563
Db 4584 TCTCTCTACCGAAAGCTTGCCTTTCTGAGATTTATCTTTGCGAGAGAGATTCGAGATTCG 4643
QY 4564 CTGAGTGTTCATCTTTTGGAGAGAGATCAATTTGATTTATTTATTTATTCATATTTCA 4623
Db 4644 CTGAGTGTTCATCTTTTGGAGAGAGATCAATTTGATTTATTTATTTATTCATATTTCA 4703

2772 AAGGAACGGGAACACCACTGTGACTGAGGGAACGAGACCTGCGGTGAGTGCACATGA 2831
2776 AAGGAACGGGAACACCACTGTGACTGAGGGAACGAGACCTGCGGTGAGTGCACATGA 2835
2832 AGGACAATCCTCATATGACAGTACTATGACCAGACATCTACGGCCCTCTCCATGAGCAGTCTGCG 2891
2836 AGGACAATCCTCATATGACAGTACTATGACCAGACATCTACGGCCCTCTCCATGAGTGCATGCG 2895
2892 TGATCCGAAAAGCCATTTCGAGAGATGTCTTTGTCAAGGGCAGCGCTGCGAGCTTCTCCGCC 2951
2896 TGATCCGAAAAGCCATTTCGAGAGATGTCTTTGTCAAGGGCAGCGCTGCGAGCTTCTCCGCC 2955
2952 GGGTCGATGACGAGCTTTTTCGGAAGATCCTTTCGGAAGCCCTATGAAATTTTTCGACAGCA 3011
2956 GGGTCGATGACGAGCTTTTTCGGAAGATCCTTTCGGAAGCCCTATGAAATTTTTCGACAGCA 3015
3012 CCCCCACAGGAGGAGTCTCAGACAGTCTTTCGAAAAGCATGAGTGAAGTTGACGCTGCGGC 3071
3016 CCCCCACAGGAGGAGTCTCAGACAGTCTTTCGAAAAGCATGAGTGAAGTTGACGCTGCGGC 3075
3072 TGGCGTTCCAGGCGGAGATGTCATCCAGAACGTTATCCGTGCTCTTCTGTGTGGGAA 3131
3076 TGGCGTTCCAGGCGGAGATGTCATCCAGAACGTTATCCGTGCTCTTCTGTGTGGGAA 3135
3132 TGAATCGAGAGATCTTCCCGTGGTCCCTGTGGCAGTGGGGCCCTTGTCATCCTCTTTT 3191
3136 TGAATCGAGAGATCTTCCCGTGGTCCCTGTGGCAGTGGGGCCCTTGTCATCCTCTTTT 3195
3192 CAGTCTGCAACATTTGTCTCCAGGCTCTGATTCGGGAGCTGAAGCGTCTGTGCACATATCA 3251
3196 CAGTCTGCAACATTTGTCTCCAGGCTCTGATTCGGGAGCTGAAGCGCTGTGCACATATCA 3255
3252 CGGAGTACCTTCTCTCCACATCCATCCATCCAGATACAGGGCCCTGSCACCATTCACG 3311
3256 CGGAGTACCTTCTCTCCACATCCATCCAGATACAGGGCCCTGSCACCATTCACG 3315
3312 CCAACAATAAAGGAGGAGATTTCTGCACAGATACAGAGAGCTCTGATGACAAACCAAG 3371
3316 CCAACAATAAAGGAGGAGATTTCTGCACAGATACAGAGAGCTCTGATGACAAACCAAG 3375
3372 CTCTCTTTTCTTTTGTAGTGTGATGTCAGTGTGCGGTGCTGCTGCGGCTGAGACCTATCA 3431
3376 CTCTCTTTTCTTTTGTAGTGTGATGTCAGTGTGCGGTGCTGCTGCGGCTGAGACCTATCA 3435
3432 GCATCGGCTCATACACACAGGAGGAGTGTGATGTCATGACAGGAGTATTCGCC 3491
3436 GCATCGGCTCATACACACAGGAGGAGTGTGATGTCATGACAGGAGTATTCGCC 3495
3492 CAGCGATGCGGAGTCTCGCATCTCTATGCTGTCCAGTTAACGGGCTGTTCAGATTCA 3551
3496 CAGCGATGCGGAGTCTCGCATCTCTATGCTGTCCAGTTAACGGGCTGTTCAGATTCA 3555
3552 CGGTGAGAGCTGAGATCTGACACAGAGCTCGATTCACCTCGGTGAGAGATCAATCACT 3611
3556 CGGTGAGAGCTGAGATCTGACACAGAGCTCGATTCACCTCGGTGAGAGATCAATCACT 3615
3612 ACATTAAGACTGTGCTCTTGAAGACCTGCGCAAGATTAAAGAACAGGCTCCGCCCTG 3671
3616 ACATTAAGACTGTGCTCTTGAAGACCTGCGCAAGATTAAAGAACAGGCTCCGCCCTG 3675
3672 ACTGGCCCCAGAGGAGGAGTGAATCTTTGAGAACGACAGAGATGAGTACCGAAGAAC 3731
3676 ACTGGCCCCAGAGGAGGAGTGAATCTTTGAGAACGACAGAGATGAGTACCGAAGAAC 3735
3732 TCCCTCTGCTCTAAGAAAGTATCTTTCACGATCAAACTAAAGAGAAAGATTGGCATTTG 3791
3736 TCCCTCTGCTCTAAGAAAGTATCTTTCACGATCAAACTAAAGAGAAAGATTGGCATTTG 3795
3792 TGGGGGCGAAGGATCGAGGAGTCTGCGTGGGAGTGGCCCTTCCGCTGCTGGTGAAGT 3851
3796 TGGGGGCGAAGGATCGAGGAGTCTGCGTGGGAGTGGCCCTTCCGCTGCTGGTGAAGT 3855

3852 TATCTGAGGCTGCATCAATGATGATGAGTGAATCAATGATATTTGGCCTTGCCGACC 3911
3856 TATCTGAGGCTGCATCAATGATGATGAGTGAATCAATGATATTTGGCCTTGCCGACC 3915
3912 TCCGAACCAACTCTCTATCATCTCTCAAGAGCCGGTGTCTTCACTGACGACATGTCAGAT 3971
3916 TCCGAACCAACTCTCTATCATCTCTCAAGAGCCGGTGTCTTCACTGACGACATGTCAGAT 3975
3972 CAATTTGGACCCCTTCAACAGTACCTGAAGACAGATGTTGGGATGCCCTGGAGAGA 4031
3976 CAATTTGGACCCCTTCAACAGTACCTGAAGACAGATGTTGGGATGCCCTGGAGAGA 4035
4032 CACACATGAAGAAGATGATGCTCAGCTACCTCGAAGACTTGAATCGAAGTGAAGAGA 4091
4036 CACACATGAAGAAGATGATGCTCAGCTACCTCGAAGACTTGAATCGAAGTGAAGAGA 4095
4092 ATGGGGATTAACCTTCTCAGTGGGGGAGACGCGAGCTCTTGTGATAGCTAGAGCCCTGCTCC 4151
4096 ATGGGGATTAACCTTCTCAGTGGGGGAGACGCGAGCTCTTGTGATAGCTAGAGCCCTGCTCC 4155
4152 GCCACGTGAAGATTTCTGATTTAGTGAAGCCACAGCTGCGCATGGACACAGACAGACT 4211
4156 GCCACGTGAAGATTTCTGATTTAGTGAAGCCACAGCTGCGCATGGACACAGACAGACT 4215
4212 TATGATTCGAAGAGACCATCCGAGAACATTTTGAGACTGTACCATGCTGACCATTTGCC 4271
4216 TATGATTCGAAGAGACCATCCGAGAACATTTTGAGACTGTACCATGCTGACCATTTGCC 4275
4272 ATGCGCTGCACAGGTTCTAGAGCTCCGATAGATTTATGTTGCTGAGCCAGGAGACAGTGG 4331
4276 ATGCGCTGCACAGGTTCTAGAGCTCCGATAGATTTATGTTGCTGAGCCAGGAGACAGTGG 4335
4332 TGGAGTTGACACCCCATTCGCTCTCTGTCACAGACAGTCCCGATTTCTATGCCATGT 4391
4336 TGGAGTTGACACCCCATTCGCTCTCTGTCACAGACAGTCCCGATTTCTATGCCATGT 4395
4392 TTTGCTGTGAGAGAACAGAGTCCGTCTCAAGGGCTGACTCTCTCTGTTGACAGAACTCT 4451
4396 TTTGCTGTGAGAGAACAGAGTCCGTCTCAAGGGCTGACTCTCTCTGTTGACAGAACTCT 4455
4452 CTTTTCTTTAGAGATTTGCCATTTCCCTGCGCTGGGGCGGGCCCTTCAATCGGCTCTCTTA 4511
4456 CTTTTCTTTAGAGATTTGCCATTTCCCTGCGCTGGGGCGGGCCCTTCAATCGGCTCTCTTA 4514
4512 CCGAACCCTTGCCCTTCTCGATTTTATCTTTGCGACAGAGTTCGGATTTGGCTGTGTTG 4571
4515 CCGAACCCTTGCCCTTCTCGATTTTATCTTTGCGACAGAGTTCGGATTTGGCTGTGTTG 4574
4572 TTTCACTTTTGAAGAGATATTTTGAATTTATTTATTTATTCATATTCATGTAACA 4631
4575 TTTCACTTTTGAAGAGATATTTTGAATTTATTTATTTATTCATATTCATGTAACA 4634
4632 AAATTTAGTTTTTGTCTTTAATTTGCACTTAAAGGTTCAAGGAAACGTTATTTAATTTG 4691
4635 AAATTTAGTTTTTGTCTTTAATTTGCACTTAAAGGTTCAAGGAAACGTTATTTAATTTG 4694
4692 TATCAGAGGCTATATGAGAGCTTTATAGCTGTATAGCTATATCTATATATTTCTGTACA 4751
4695 TATCAGAGGCTATATGAGAGCTTTATAGCTGTATAGCTATATCTATATATTTCTGTACA 4754
4752 TAGCCTATATTTACAGTAAGTAAGTGTATTTATTTATTTAAATAAGCACTGTGC 4811
4755 TAGCCTATATTTACAGTAAGTAAGTGTATTTATTTATTTAAATAAGCACTGTGC 4814
4812 TAAAAAAA 4819
4815 TAAATACA 4822

RESULT 5
AR091520 4781 bp DNA PAT 07-SEP-2000
LOCUS AR091520 4781 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5994130.

```
|||||
Db 556 GCACAAAGCTGAGAGACTGGGCAAGAAAGCTGAATGAAGTTGGCCAGACGCTGTT 615
612 CCTGCGAAGGGTGTGTGATCTTCTGCCACACAGGCTCATCTGTCATCGTGTGC 671
Db 616 CCGTCGGAAGGGTGTGTGATCTTCTGCCACACAGGCTCATCTGTCATCGTGTGC 675
672 TGATGATCAGCAGCTGGCTGGCTTCACTGAGACAGCCTTCATGGTGAACACCTCTGG 731
Db 676 TGATGATCAGCAGCTGGCTGGCTTCACTGAGACAGCCTTCATGGTGAACACCTCTGG 735
732 AGTATACCCAGGCAAGAGTCTAACCCTGACGTACAGTGTGTGTATGCTGGCCCTCC 791
736 AGTATACCCAGGCAAGAGTCTAACCCTGACGTACAGTGTGTGTATGCTGGCCCTCC 795
792 TCCGACGGAATATGCTGGCTGTGCTGCTGACCTGACCTGGGCAATGAAATACCAA 851
Db 796 TCCGACGGAATATGCTGGCTGTGCTGCTGACCTGACCTGGGCAATGAAATACCAA 855
852 CCGGTGCTCGCTTGGGGGGCCATCCTAACCATGGCATTTAAGAAGATCCTTAAGTTAA 911
Db 856 CCGGTGCTCGCTTGGGGGGCCATCCTAACCATGGCATTTAAGAAGATCCTTAAGTTAA 915
912 AGAATATTAAAGAAATCCCTGGGTGACCTCATCAATTTGCTCAACGATGGGCA 971
916 AGAATATTAAAGAAATCCCTGGGTGACCTCATCAATTTGCTCAACGATGGGCA 975
972 GAATGTTGAGGACGACCCCTGGACGCTGCTGGAGAGACCCGTTGTCATCT 1031
Db 976 GAATGTTGAGGACGACCCCTGGACGCTGCTGGAGAGACCCGTTGTCATCT 1035
1032 TAGCATGATTTATATGATTAATTTCTGGAGCAACAGGCTTCTGGATCAGCTGTT 1091
Db 1036 TAGCATGATTTATATGATTAATTTCTGGAGCAACAGGCTTCTGGATCAGCTGTT 1095
1092 TTATCTCTTTTACCGACATGATGTTGTCATCAGGCTCAGACATATTTCAAGAGAA 1151
1096 TTATCTCTTTTACCGACATGATGTTGTCATCAGGCTCAGACATATTTCAAGAGAA 1155
1152 AATGGTGGCCGCGCAGGATACGCTGCCAAGAAAGTAAGTAATGCTTCTTCAATTA 1211
Db 1156 AATGGTGGCCGCGCAGGATACGCTGCCAAGAAAGTAAGTAATGCTTCTTCAATTA 1215
1212 AATTTATCAAAATGATGCTCGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAATCCGCG 1271
Db 1216 AATTTATCAAAATGATGCTCGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAATCCGCG 1275
1272 AGAAGAGCGTGGATATTTGAAAAAGCCGGTACTTCCAGACATCACTGGGTGG 1331
Db 1276 AGAAGAGCGTGGATATTTGAAAAAGCCGGTACTTCCAGACATCACTGGGTGG 1335
1332 CTCCTCATGTTGGTGTGATGCTCCAGCGGTGGTGAACCTTCTGTTATATGACCCCTGGCT 1391
Db 1336 CTCCTCATGTTGGTGTGATGCTCCAGCGGTGGTGAACCTTCTGTTATATGACCCCTGGCT 1395
1392 TCGATCTGACAGCAGACAGGCTTTCACAGTGTGACAGTCTTCAATTCATGACTTTTG 1451
Db 1396 TCGATCTGACAGCAGACAGGCTTTCACAGTGTGACAGTCTTCAATTCATGACTTTTG 1455
1452 CTTTGAAGATACACCGTTTCACTAAAGTCCCTCAGAAAGCTCAGTGGCTGTTGACA 1511
Db 1456 CTTTGAAGATACACCGTTTCACTAAAGTCCCTCAGAAAGCTCAGTGGCTGTTGACA 1515
1512 GATTTAAGAGTTGTTGTTATGAAAGAGTTCACATGATTAAGAAACAAACAGCAGATC 1571
Db 1516 GATTTAAGAGTTGTTGTTATGAAAGAGTTCACATGATTAAGAAACAAACAGCAGATC 1575
1572 CTCACATCAAGATAGAGATGAAAAATGCACCTGGGATGGGATCTCCACACTCAGATA 1631
Db 1576 CTCACATCAAGATAGAGATGAAAAATGCACCTGGGATGGGATCTCCACACTCAGATA 1635
1632 TCCAGAACTGCGCCCAAGCTGACCCCAAAATGAAAAAGACAAAGAGGCTTCCAGGGGCA 1691
|||||

Db 1636 TCCAGAACTGCGCCCAAGCTGACCCCAAAATGAAAAAGACAAAGAGGCTTCCAGGGGCA 1695
1692 AGAAAGAGAGGTGAGGAGCTGCAGCGCCTGACATCAGCGGTGCTGGCAGACAGA 1751
Db 1696 AGAAAGAGAGGTGAGGAGCTGCAGCGCCTGACATCAGCGGTGCTGGCAGACAGA 1755
1752 AAGGCCACCTCCTCTGACAGTACAGAGCGGCCCAAGTCCGAGAGAAAGAACCAAGC 1811
Db 1756 AAGGCCACCTCCTCTGACAGTACAGAGCGGCCCAAGTCCGAGAGAAAGAACCAAGC 1815
1812 ACATCCACTGGGGCCACCTGGCTTACAGAGGACACTGACAGCATGCTGAGATATC 1871
Db 1816 ACATCCACTGGGGCCACCTGGCTTACAGAGGACACTGACAGCATGCTGAGATATC 1875
1872 AAGAGGTAAATGTTGGAATCTGCGGCAAGTGTGGAAAGTGAAGAAACCTTCTCATATT 1931
Db 1876 AAGAGGTAAATGTTGGAATCTGCGGCAAGTGTGGAAAGTGAAGAAACCTTCTCATATT 1935
1932 CAGCCATTTTAAAGCAGATGACGCTTCTAGAGGCGACGATTCGAATCAGTGAACCTTCG 1991
Db 1936 CAGCCATTTTAAAGCAGATGACGCTTCTAGAGGCGACGATTCGAATCAGTGAACCTTCG 1995
1992 CTTATGTGGCCAGAGCGCTGGATCTCAATGCTACCTGAGAGACAAACATCCTGTTTG 2051
Db 1996 CTTATGTGGCCAGAGCGCTGGATCTCAATGCTACCTGAGAGACAAACATCCTGTTTG 2055
2052 GGAAGGAATATGATGAAGAAAGATCAACTCTGTGCTGACAGCTGCTGAGAGGCTG 2111
Db 2056 GGAAGGAATATGATGAAGAAAGATCAACTCTGTGCTGACAGCTGCTGAGAGGCTG 2115
2112 ACCTGGCATTTCTCCAGACGACGCTGACGAGATTTGAGAGCGGAGGCAACCTGA 2171
Db 2116 ACCTGGCATTTCTCCAGACGACGCTGACGAGATTTGAGAGCGGAGGCAACCTGA 2175
2172 GCGGTGGGAGCGCCAGAGGATCAGCTTGGCCGGGCTGTGATGATGACAGAGCATCT 2231
Db 2176 GCGGTGGGAGCGCCAGAGGATCAGCTTGGCCGGGCTGTGATGATGACAGAGCATCT 2235
2232 ACATCTGTGACGACCCCTCAGTGTGCTTATGATGCTTGGGCAACCATCTTCAATA 2291
Db 2236 ACATCTGTGACGACCCCTCAGTGTGCTTATGATGCTTGGGCAACCATCTTCAATA 2295
2292 GTGCTATCCGGAACATCTCAAGTCCAAGACAGTCTGTTGTTACCAACAGTTACGT 2351
Db 2296 GTGCTATCCGGAACATCTCAAGTCCAAGACAGTCTGTTGTTACCAACAGTTACGT 2355
2352 ACCTGTTGACCTGTATGAAGATCTTCATGAAGAGGCGTGTATTCAGGAAAGAGCA 2411
Db 2356 ACCTGTTGACCTGTATGAAGATCTTCATGAAGAGGCGTGTATTCAGGAAAGAGCA 2415
2412 CCCATGAGAACTGATGAATTTAAATGGTGAATGCTATGCTACATTTTAAATACCTGTTGC 2471
Db 2416 CCCATGAGAACTGATGAATTTAAATGGTGAATGCTATGCTATGCTATTTTAAATACCTGTTGC 2475
2472 TGGGAGAGACACCGCAGTGTGATCAATTTCAAAAAGAAACAGTGGTTCACAGAA 2531
Db 2476 TGGGAGAGACACCGCAGTGTGATCAATTTCAAAAAGAAACAGTGGTTCACAGAA 2535
2532 AGTCACAAGACAGGCTCTTAAACAGATCAATTAAGAAAGAAAAACAGTAAGCCAG 2591
Db 2536 AGTCACAAGACAGGCTCTTAAACAGATCAATTAAGAAAGAAAAACAGTAAGCCAG 2595
2592 AGGAAGGAGAGCTGTGACGTGGAAGAAAGGAGGAGGTTCACTGGCCCTGCTGATAT 2651
Db 2596 AGGAAGGAGAGCTGTGACGTGGAAGAAAGGAGGAGGTTCACTGGCCCTGCTGATAT 2655
2652 ATGCTGTACATCCAGGCTGCTGGGGGCCCTTGGCATTTCTGTTATATGAGCCCTTT 2711
Db 2656 ATGCTGTACATCCAGGCTGCTGGGGGCCCTTGGCATTTCTGTTATATGAGCCCTTT 2715
2712 TCATGCTGAATGTAGGACAGACCGCTTCAGACCTGCTGTGATTAATGATCAAGC 2771
Db 2716 TCATGCTGAATGTAGGACAGACCGCTTCAGACCTGCTGTGATTAATGATCAAGC 2775
```



```
|||||
Db 1861 CATTTCGACATTTTAGCCAGATGACGCTTCTAGAGGAGACATTCATCTAGTGAAC 1920
|||
Qy 1987 CTTTCGCTTATGTGGCCCGACGAGCGCTGGATCTCTCAATGATCTACTCTGAGAGACAATCTCT 2046
|||
Db 1921 CTTTCGCTTATGTGGCCCGACGAGCGCTGGATCTCTCAATGATCTACTCTGAGAGACAATCTCT 1980
|||
Qy 2047 GTTTGGGAAGGAATATGATGAAGAAAGATACAACTCTGTCTGAAACAGCTGTGCTGAG 2106
|||
Db 1981 GTTTGGGAAGGAATATGATGAAGAAAGATACAACTCTGTCTGAAACAGCTGTGCTGAG 2040
|||
Qy 2107 GCGTGCACCTGGCATTCTCTCCAGACGACGACCTGACGAGATTTGGAAGGAGAGGCCAA 2166
|||
Db 2041 GCGTGCACCTGGCATTCTCTCCAGACGACGACCTGACGAGATTTGGAAGGAGAGGCCAA 2100
|||
Qy 2167 CCTGAGCGGTGGGCGACGCCAGAGAGATCAGCCCTTGCCGGGCTTGTATATGTGACAGAG 2226
|||
Db 2101 CCTGAGCGGTGGGCGACGCCAGAGAGATCAGCCCTTGCCGGGCTTGTATATGTGACAGAG 2160
|||
Qy 2227 CATCTCATCTGGAGACGCCCTGACGTGCTTAGATGCCATGTGGGCAACCATCTT 2286
|||
Db 2161 CATCTCATCTGGAGACGCCCTGACGTGCTTAGATGCCATGTGGGCAACCATCTT 2220
|||
Qy 2287 CAATAGTCTATCTCGGAACATCTCAAGTCCAAAGACAGTTCTGTTGTACCCACAGTT 2346
|||
Db 2221 CAATAGTCTATCTCGGAACATCTCAAGTCCAAAGACAGTTCTGTTGTACCCACAGTT 2280
|||
Qy 2347 ACAGTACCTGTGTGACTGTGTATGATGATCTTCATGAAAGAGGCGTGTATACGGAAG 2406
|||
Db 2281 ACAGTACCTGTGTGACTGTGTATGATGATCTTCATGAAAGAGGCGTGTATACGGAAG 2340
|||
Qy 2407 AGGACACCATGAGGAATGTGATTAATTAATGATGATGCTACATTTTAAATACCT 2466
|||
Db 2341 AGGACACCATGAGGAATGTGATTAATTAATGATGATGCTACATTTTAAATACCT 2400
|||
Qy 2467 GTTTCGTGGAGAGACACCGCCAGTTGAGATCAATTCAAAAAGAAACCAAGTGTTCACA 2526
|||
Db 2401 GTTTCGTGGAGAGACACCGCCAGTTGAGATCAATTCAAAAAGAAACCAAGTGTTCACA 2460
|||
Qy 2527 GAAGAACTCAACAGACAAAGGCTCTTAACAGAGATCAATTAAGAAAGAAAAAGAGATAA 2586
|||
Db 2461 GAAGAACTCAACAGACAAAGGCTCTTAACAGAGATCAATTAAGAAAGAAAAAGAGATAA 2520
|||
Qy 2587 GCCAGAGGAAGGCGAGCTTGTGACGTGGAAGAGAAAGGCGGTTCAAGGCCCTGTGC 2646
|||
Db 2521 GCCAGAGGAAGGCGAGCTTGTGACGTGGAAGAGAAAGGCGGTTCAAGGCCCTGTGC 2580
|||
Qy 2647 AGTATATGTGTCTACATCCAGGCTGCTGGGGGCCCTTGGCATTCCTGTTATATGAC 2706
|||
Db 2581 AGTATATGTGTCTACATCCAGGCTGCTGGGGGCCCTTGGCATTCCTGTTATATGAC 2640
|||
Qy 2707 CTTTTATGCTGAATGATAGGACACACCGCTTTCAGACACCTGGTGTGATCTGTGAT 2766
|||
Db 2641 CTTTTATGCTGAATGATGAGGACACACCGCTTTCAGACACCTGGTGTGATCTGTGAT 2700
|||
Qy 2767 CAAGCAAGGAAGCGGAACACCACTGTGACTCGAGGGAAGAGACCTCGGAGTGTGACAG 2826
|||
Db 2701 CAAGCAAGGAAGCGGGAACACCACTGTGACTCGAGGGAAGAGACCTCGGAGTGTGACAG 2760
|||
Qy 2827 CAGTAAGGACAATCTCATATGACGATATGCCAGCATCTACGCCCTCTCCATGGCAGT 2886
|||
Db 2761 CAGTAAGGACAATCTCATATGACGATATGCCAGCATCTACGCCCTCTCCATGGCAGT 2820
|||
Qy 2887 CAGTCTATCTGAAAGCCATGTCAGAGAGTGTCTTGTGTAAGGGCACGCTGCACAGCTTC 2946
|||
Db 2821 CAGTCTATCTGAAAGCCATGTCAGAGAGTGTCTTGTGTAAGGGCACGCTGCACAGCTTC 2880
|||
Qy 2947 CTCGCGGCTCATAGCAGGCTTTTCCGAAGGATCTTCGAAGCCCTTGAAGTTTGTGA 3006
|||
Db 2881 CTCGCGGCTCATAGCAGGCTTTTCCGAAGGATCTTCGAAGCCCTTGAAGTTTGTGA 2940
|||
Qy 3007 CAGACCCCCACAGGAGATCTCAACAAGTTTCCAAAGACATGATGAAGTTGACGT 3066
|||
|||||
Db 2941 CAGACCCCCACAGGAGATCTCAACAAGTTTCCAAAGACATGATGAAGTTGACGT 3000
|||
Qy 3067 GCGGCTGCCCTTCCAGGCCAGAGATTTTCAATCCAGAACGTTATCTGCTGTCTGTGT 3126
|||
Db 3001 GCGGCTGCCCTTCCAGGCCAGAGATTTTCAATCCAGAACGTTATCTGCTGTCTGTGT 3060
|||
Qy 3127 GGGAAATATGACAGGAGTCTTCCCGTGTCTTGTGGAGTGGGGCCCTTGTCAATCT 3186
|||
Db 3061 GGGAAATATGACAGGAGTCTTCCCGTGTCTTGTGGAGTGGGGCCCTTGTCAATCT 3120
|||
Qy 3187 CTTTTAGTCTGACATTTGCTTCCAGAGGCTCTGATTCGGGAGCTGAAGGCTGTGACAA 3246
|||
Db 3121 CTTTTAGTCTGACATTTGCTTCCAGAGGCTCTGATTCGGGAGCTGAAGGCTGTGACAA 3180
|||
Qy 3247 TATCAGCAGTCACTTCTCTTCCACATCAGGTCAGCATACAGGCTTGTGCACAT 3306
|||
Db 3181 TATCAGCAGTCACTTCTCTTCCACATCAGGTCAGCATACAGGCTTGTGCACAT 3240
|||
Qy 3307 CCAAGCCTACATTAAGGGAGAGTTTCTGCACAGATACAGAGGCTGTGATGACAA 3366
|||
Db 3241 CCAAGCCTACATTAAGGGAGAGTTTCTGCACAGATACAGAGGCTGTGATGACAA 3300
|||
Qy 3367 CCAAGCTCTTTTTTTTTTTTGTATACGTGTGCGATGCGGTGCGTGGCTGACCT 3426
|||
Db 3301 CCAAGCTCTTTTTTTTTTTTGTATACGTGTGCGATGCGGTGCGTGGCTGACCT 3360
|||
Qy 3427 CATCAGATGGCCCTCATTCACACACACAGGGGCTGATATGTTCTTATGACGGGCGAT 3486
|||
Db 3361 CATCAGATGGCCCTCATTCACACACAGGGGCTGATATGTTCTTATGACGGGCGAT 3420
|||
Qy 3487 TCCCCAGCCTATGCGGGGTGCGCATCTCTATGCTGTCAGTTAAGGGGCTGTTCCA 3546
|||
Db 3421 TCCCCAGCCTATGCGGGGTGCGCATCTCTATGCTGTCAGTTAAGGGGCTGTTCCA 3480
|||
Qy 3547 GTTTACGCTCAGACTGACATCTGAGACAGAGCTGATTCACCTCGGTGAGAGATCAA 3606
|||
Db 3481 GTTTACGCTCAGACTGACATCTGAGACAGAGCTGATTCACCTCGGTGAGAGATCAA 3540
|||
Qy 3607 TCACCTATTAAGACTGTGTCTTGGAAAGCACCTGGAGATTAAGAAAGGCTTCCCTC 3666
|||
Db 3541 TCACCTATTAAGACTGTGTCTTGGAAAGCACCTGGAGATTAAGAAAGGCTTCCCTC 3600
|||
Qy 3667 CCGTCACTGGCCCCAGGAGGAGAGAGTGTGACCTTGGAAAGGAGAGATGATACGAGA 3726
|||
Db 3601 CCGTCACTGGCCCCAGGAGGAGAGAGTGTGACCTTGGAAAGGAGAGATGATACGAGA 3660
|||
Qy 3727 AAACCTCCTCTGCTCTTAAGAAAGTATCTTCAGATCAAACTTAAGAAAGATGG 3786
|||
Db 3661 AAACCTCCTCTGCTCTTAAGAAAGTATCTTCAGATCAAACTTAAGAAAGATGG 3720
|||
Qy 3787 CATTTGGGGGGGACAGAGATCAGGAGATCCTGCTGGGGATGGCCCTCTTCGCTGGT 3846
|||
Db 3721 CATTTGGGGGGGACAGAGATCAGGAGATCCTGCTGGGGATGGCCCTCTTCGCTGGT 3780
|||
Qy 3847 GGAATTAATCTGAGAGCTGATCAAGATTAAGATGAGATGAGATCAATCTATGACCTTGC 3906
|||
Db 3781 GGAATTAATCTGAGAGCTGATCAAGATTAAGATGAGATGAGATCAATCTATGACCTTGC 3840
|||
Qy 3907 CGACCTCCGAAGCAACTCTCTATCATTTCTCAAGACCGGCTGTCTACAGTGGCAGT 3966
|||
Db 3841 CGACCTCCGAAGCAACTCTCTATCATTTCTCAAGACCGGCTGTCTACAGTGGCAGT 3900
|||
Qy 3967 CAGATCAAAATTTGGAGCCCTTCAACACAGTACACTGAAGACAGATTTGGATGGCCCTGGA 4026
|||
Db 3901 CAGATCAAAATTTGGAGCCCTTCAACACAGTACACTGAAGACAGATTTGGATGGCCCTGGA 3960
|||
Qy 4027 GAGGACACACATGAAGAAATGATTTGCTCAGCTACTGTGAACCTGAAATCTGAAGTAT 4086
|||
Db 3961 GAGGACACACATGAAGAAATGATTTGCTCAGCTACTGTGAACCTGAAATCTGAAGTAT 4020
|||
Qy 4087 GGAAGATGGGGATTAATCTTCACTGAGTGGGGGAACGGCAGCTCTTGTGCTACTAGAGCCT 4146
|||
Db 4021 GGAAGATGGGGATTAATCTTCACTGAGTGGGGGAACGGCAGCTCTTGTGCTACTAGAGCCT 4080
|||
```

ACCESSION AR091520
VERSION AR091520.1 GI:10018275
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 4781)
TITLE Shyjan, A.
JOURNAL Multidrug resistance-associated polypeptide
FEATURES Patent: US 5994130-A 1 30-NOV-1999;
Location/Qualifiers
Source 1..4781
/organism="unknown"
BASE COUNT 1217 a 1161 c 1232 g 1165 t 6 others
ORIGIN

Query Match 98.5%; Score 4775.4; DB 9; Length 4781;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4773; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

67 TGATGTGAACATACAGCTGTGAGCCCTCGAACCTCCACTCAGAGAGATGAAGATAT 126
1 TGATGTGAACATACAGCTGTGAGCCCTCGAACCTCCACTCAGAGAGATGAAGATAT 60
127 CGACATAGAAAGAGATATCATCCAGCTCTGGGTATGAAAGTGAAGAGAGAAC 186
61 CGACATAGAAAGAGATATCATCCAGCTCTGGGTATGAAAGTGAAGAGAGAAC 120
187 CAGACTCTTGGGAGCAGACAGACCGTGAAGATTCCAAATTCCAGAGACTCGACGTT 246
121 CAGACTCTTGGGAGCAGACAGACCGTGAAGATTCCAAATTCCAGAGACTCGACGTT 180
247 GGAATGCCAAGATGCTTGAAGACAGCCGAGCCGAGGGGCTCTCTTGAATGCTC 306
181 GGAATGCCAAGATGCTTGAAGACAGCCGAGCCGAGGGGCTCTCTTGAATGCTC 240
307 CATGCATCTCAGCTCAGAAATCTGGATGAGAGACATCCCAAGGGAAAGTACCATATG 366
241 CATGCATCTCAGCTCAGAAATCTGGATGAGAGACATCCCAAGGGAAAGTACCATATG 300
367 CTGAGTCTCTGAAGCCATCCGAGTACTTGCACACACAGCCCAAGTGGACATGC 426
301 CTGAGTCTCTGAAGCCATCCGAGTACTTGCACACACAGCCCAAGTGGACATGC 360
427 TGGGCTTTTTCCTGATGACTTTTCTGGGCTTCTCTGCGCCGCTGGGCCACAA 486
361 TGGGCTTTTTCCTGATGACTTTTCTGGGCTTCTCTGCGCCGCTGGGCCACAA 420
487 GAAAGGGAGCTCTCAATGAGAGAGTGTCTCTCTGCAAGACGAGTCTTGTGAGCT 546
421 GAAAGGGAGCTCTCAATGAGAGAGTGTGTCTCTGCAAGACGAGTCTTGTGAGCT 480
547 GAACCTGCAAGAGCTAGAGAGACTGTGGCAAGAGAGCTGAATGAAGTTGGGCCAGACC 606
481 GAACCTGCAAGAGCTAGAGAGACTGTGGCAAGAGAGCTGAATGAAGTTGGGCCAGACC 540
607 TGGTCCCTGGAGAGGCTGTGGATCTCTGCGCCAGCAGGCTCACTCTGCATGCT 666
541 TGGTCCCTGGAGAGGCTGTGGATCTCTGCGCCAGCAGGCTCACTCTGCATGCT 600
667 GTGCTGATGATCAAGCAGCTGTGCTGCTCAAGTGAAGCAGCTCATGTAAGACACT 726
601 GTGCTGATGATCAAGCAGCTGTGCTGCTCAAGTGAAGCAGCTCATGTAAGACACT 660
727 CTGAGACTATACAGGAGCAGAGCTTAACCTGCAGTACAGCTTGTGTAAGTCTGGG 786
661 CTGAGACTATACAGGAGCAGAGCTTAACCTGCAGTACAGCTTGTGTAAGTCTGGG 720
787 CCTCTCTGAGGAAATCGTGGCGTCTTGGTGGCTGACTGACTTGGCATTTGAATTA 846
721 CCTCTCTGAGGAAATCGTGGCGTCTTGGTGGCTGACTGACTTGGCATTTGAATTA 780

847 CCGAACCGGTGCTCCGCTTGGGGGGGCCATCTTAACATGGCATTTTAAGAAATCTTAA 906
781 CCGAACCGGTGCTCCGCTTGGGGGGGCCATCTTAACATGGCATTTTAAGAAATCTTAA 840
907 GTTAAAGAACTTAAAGAAATCCCTGGGTAGCTCATCAACATTTCTCCAAAGATGG 966
841 GTTAAAGAACTTAAAGAAATCCCTGGGTAGCTCATCAACATTTCTCCAAAGATGG 900
967 GCAGAAATGTTTGAAGAGCAGCCGTTGGCAGCTGCTGGCTGAGAGACCGTGTGTC 1026
901 GCAGAAATGTTTGAAGAGCAGCCGTTGGCAGCTGCTGGCTGAGAGACCGTGTGTC 960
1027 CATCTTAGCATGATTTATTAATGTAATTTCTGGACCAACAGGCTTCTGGATCAGC 1086
961 CATCTTAGCATGATTTATTAATGTAATTTCTGGACCAACAGGCTTCTGGATCAGC 1020
1087 TGTTTTATCCCTTTTACCCAGCAATGATGTTTCATCAGCGCTCAGACATATTAG 1146
1021 TGTTTTATCCCTTTTACCCAGCAATGATGTTTCATCAGCGCTCAGACATATTAG 1080
1147 GAGAAATGCGTGGCCGACGAGATGAACGTGTCCAGAAAGTGAAGTCTTACTTTA 1206
1081 GAGAAATGCGTGGCCGACGAGATGAACGTGTCCAGAAAGTGAAGTCTTACTTTA 1140
1207 CATTAATTTTAAATAATGTAATGCTTGGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAAT 1266
1141 CATTAATTTTAAATAATGTAATGCTTGGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAAT 1200
1267 CCGGAGAGAGAGCTCGGATTTTGGAAAAAGCCGGGTACTTCCAGACATCAGCTGTGG 1326
1201 CCGGAGAGAGAGCTCGGATTTTGGAAAAAGCCGGGTACTTCCAGACATCAGCTGTGG 1260
1327 TGTGCTCCCATTTGGTGGTATGATGCGAGCGTGTGACCTTCTGTTCATATGACCT 1386
1261 TGTGCTCCCATTTGGTGGTATGATGCGAGCTGTGACCTTCTGTTCATATGACCT 1320
1387 GGGCTTCATCTGACAGACGACAGCGTTTTCACAGTGTGACAGTCTTCAATTCATGAC 1446
1321 GGGCTTCATCTGACAGACGACAGCGTTTTCACAGTGTGACAGTCTTCAATTCATGAC 1380
1447 TTTTGCCTTGAAGTAACCGTTTTCAGTAAGTCCCTCAGAAAGCTCAGTGGCTG 1506
1381 TTTTGCCTTGAAGTAACCGTTTTCAGTAAGTCCCTCAGTAAGTCCCTCAGTAAGTCCCT 1440
1507 TGACAGATTAAAGATTTGTTCTTAATGAAAGAGTTCACATGATAAAGCAAAACAGC 1566
1441 TGACAGATTAAAGATTTGTTCTTAATGAAAGAGTTCACATGATAAAGCAAAACAGC 1500
1567 CAGTCTCATCATCAAGATGAGATGAAAAATGCACTTGGCATGGAGCTCCTCCACTC 1626
1501 CAGTCTCATCATCAAGATGAGATGAAAAATGCACTTGGCATGGAGCTCCTCCACTC 1560
1627 CAGTATCCAGAACTGCCCCAGCTACCCCCAAATGAAAAAGCAAGAGGCTTCCAG 1686
1561 CAGTATCCAGAACTGCCCCAGCTACCCCCAAATGAAAAAGCAAGAGGCTTCCAG 1620
1687 GGGCAAGAAAGAGAGTGAAGCAGCTGAGGAGCACTGAGCACTCAGGCGGTGCTGGCAGA 1746
1621 GGGCAAGAAAGAGAGTGAAGCAGCTGAGGAGCACTGAGCACTCAGGCGGTGCTGGCAGA 1680
1747 GCAGAAAGCCACCTCTCTGAGACAGTGAAGCGGCCCAAGTCCCGAAGAAAGAAAG 1806
1681 GCAGAAAGCCACCTCTCTGAGACAGTGAAGCGGCCCAAGTCCCGAAGAAAGAAAG 1740
1807 CAAGCAATCCACCTGGGACACCTGCGCTTACAGAGACACTGCACAGCATGATCTGGA 1866
1741 CAAGCAATCCACCTGGGACACCTGCGCTTACAGAGACACTGCACAGCATGATCTGGA 1800
1867 GATCCAAAGAGGTTAACTGGTTGAAATCTGGGCGCAGTGGGAATGGAAAAAAGCTCTCT 1926
1801 GATCCAAAGAGGTTAACTGGTTGAAATCTGGGCGCAGTGGGAATGGAAAAAAGCTCTCT 1860
1927 CATTTACAGCCATTTTAGGCCAGATGACGCTTCTAGAGGGCAGCATTTGCAATCAGTGAAC 1986

| | | | |
|----|------|--|------|
| QY | 424 | TGAGGGCTTTTTCGTAAGACATTTTGTGAGCTTCTCTCTGACCGGTGGCCCA | 483 |
| Db | 361 | TGCTGGCTTTTTCGTATGACTTTTGTGAGCTTCTCTGACCGGTGGCCCA | 420 |
| QY | 484 | CAAGAAGGGAGCTCTCAATGGAAGACGTGGTCTCTTCCAAAGCAGACTTCTGA | 543 |
| Db | 421 | CAAGAAGGGAGCTCTCAATGGAAGACGTGGTCTCTTCCAAAGCAGACTTCTGA | 480 |
| QY | 544 | CGTAAGCTCAGAAAGACTAGAGACTGTGGCAGAAAGACTGAATGAATTGGGCCAGA | 603 |
| Db | 481 | CGTAAGCTCAGAGAGACTAGAGAGACTGTGGCAAGAAAGACTGAATGAATTGGGCCAGA | 540 |
| QY | 604 | CGCTGCTTCCCTGGGAAGGGTGTGTGGATCTTCTGGCCAGCAGAGCTACTCTGTGCAT | 663 |
| Db | 541 | CGCTGCTTCCCTGGGAAGGGTGTGTGGATCTTCTGGCCAGCAGAGCTACTCTGTGCAT | 600 |
| QY | 664 | CGTGTGCTGTATGATCAACGACAGCTGGCTTCAGTGAACACAGCCCTCATGGTGAACA | 723 |
| Db | 601 | CGTGTGCTGTATGATCAACGACAGCTGGCTTCAGTGAACACAGCCCTCATGGTGAACA | 660 |
| QY | 724 | CCCTTGGAGTATACCAAGCAAGCTCTAACCTGACAGTACAGCTTGTGTAGTGCT | 783 |
| Db | 661 | CCCTTGGAGTATACCAAGCAAGCTCTAACCTGACAGTACAGCTTGTGTAGTGCT | 720 |
| QY | 784 | GGGGCTCTCTGTGACGGAAATCGTGGCGTCTGGTGGCTTGACACTGACTGGGCATTGGA | 843 |
| Db | 721 | GGGGCTCTCTGTGACGGAAATCGTGGCGTCTGGTGGCTTGACACTGACTGGGCATTGGA | 780 |
| QY | 844 | TTACCGAACCGGTGTCGGCTTGGCGGGGGCCATCTAACCATGGCATTTAAGAAATCCT | 903 |
| Db | 781 | TTACCGAACCGGTGTCGGCTTGGCGGGGGCCATCTAACCATGGCATTTAAGAAATCCT | 840 |
| QY | 904 | TAACTTAAAGAACATTTAAAGAAATCCTTGGGTGAAGCTCATCAACTTTGCTCAACGA | 963 |
| Db | 841 | TAACTTAAAGAACATTTAAAGAAATCCTTGGGTGAAGCTCATCAACTTTGCTCAACGA | 900 |
| QY | 964 | TGGGAGAGATGTGTTGGAGAGACGCGTTGGAGAGCTGTGGCTGGAGAGCCCGTTGT | 1022 |
| Db | 901 | TGGGAGAGATGTGTTGGAGAGAGCCGTTGGAGAGCTGTGGCTGGAGAGCCCGTTGT | 960 |
| QY | 1024 | TGCCATCTTAGCATGATTTATATATATATCTTGGGACCAACAGCTTCTGGGATC | 1088 |
| Db | 961 | TGCCATCTTAGCATGATTTATATATATATCTTGGGACCAACAGCTTCTGGGATC | 1022 |
| QY | 1084 | AGCTGTATTATCTCTTTTACCAGAAATGATTTTGCAATCAGGCTCACAGCATATTT | 1144 |
| Db | 1021 | AGCTGTATTATCTCTTTTACCAGAAATGATTTTGCAATCAGGCTCACAGCATATTT | 1088 |
| QY | 1144 | CAGGAGAAATGCGTGGGCGCCAGAGATGAACGTGCAGAAAGTGAATGAAGTCTTAC | 1202 |
| Db | 1081 | CAGGAGAAATGCGTGGGCGCCAGAGATGAACGTGCAGAAAGTGAATGAAGTCTTAC | 1144 |
| QY | 1204 | TTACATTTAAATTTATCAAAATGTATGCTGGCTGAAGCAATTTTCTCAGAGTTCGAA | 1266 |
| Db | 1141 | TTACATTTAAATTTATCAAAATGTATGCTGGCTGAAGCAATTTTCTCAGAGTTCGAA | 1202 |
| QY | 1264 | AATCCGAGAGAGAGAGCTGGGAATTTGGAAAAAGCCGGGTAATCTTCAGAGCATCACTGT | 1322 |
| Db | 1201 | AATCCGAGAGAGAGAGCTGGGAATTTGGAAAAAGCCGGGTAATCTTCAGAGCATCACTGT | 1266 |
| QY | 1324 | GGGTGTGCTCCCATTTGTGTGTGATTTGCCAGCGGTGTGACCTTCTCTTATATGAC | 1388 |
| Db | 1261 | GGGTGTGCTCCCATTTGTGTGTGATTTGCCAGCGGTGTGACCTTCTCTTATATGAC | 1322 |
| QY | 1384 | CTGTGGCTTGCATCTGCACGACGACAGGCTTTCACAGTGGTGCAGTCTTCAATTCAT | 1444 |
| Db | 1321 | CTGTGGCTTGCATCTGCACGACGACAGGCTTTCACAGTGGTGCAGTCTTCAATTCAT | 1388 |
| QY | 1444 | GACTTGTCTTTGAAGTAACCGGTTTACGTAAATGCTCCCTTCACAAAGCTCAGTGGC | 1502 |
| Db | 1381 | GACTTGTCTTTGAAGTAACCGGTTTACGTAAATGCTCCCTTCACAAAGCTCAGTGGC | 1444 |

| | | | |
|----|------|---|-------|
| QY | 1504 | TGTTGACAGATTTTAAGAGTTTGTTTTCATATGGAAGAGGTCCATGATTAAGAACAAC | 1563 |
| Db | 1441 | TGTTTGACAGATTTTAAGAGTTTGTTTTCATATGGAAGAGGTTCACATGATTAAGAACAAC | 1500 |
| QY | 1564 | AGCCAGTCTCACATCAATCAAGTATAGATTAAGAAAAATGCACTTGGCATGGGACTCTCCCA | 16233 |
| Db | 1501 | AGCCAGTCTCACATCAAGTATAGATTAAGAAAAATGCACTTGGCATGGGACTCTCCCA | 1560 |
| QY | 1624 | CTCCAGTATCCAGAACTCGCCCAAGCTGACCCCCCAAAATGAAAAAGACAAAGAGGGCTTC | 1663 |
| Db | 1561 | CTCCAGTATCCAGAACTCGCCCAAGCTGACCCCCCAAAATGAAAAAGACAAAGAGGGCTTC | 1620 |
| QY | 1684 | CAGGGCGAAGAAAGAGAGGTGAGGCGAGCTGCAGGCGCATGTGAGCATCAGGGGGTCTG96C | 17433 |
| Db | 1621 | CAGGGCGAAGAAAGAGAGGTGAGGCGAGCTGCAGGCGCATGTGAGCATCAGGGGGTCTG96C | 1680 |
| QY | 1744 | AGACACAAAGAGGCACTCTCTCTGACAGTGCAGAGGGGGCCACATCCCGAAGAGAGAA | 18033 |
| Db | 1681 | AGACACAAAGAGGCACTCTCTCTGACAGTGCAGAGGGGGCCACATCCCGAAGAGAGAA | 1740 |
| QY | 1804 | AGGCAACACATCCACTCTGGGCGCCACTGCGCTTACAGAGACATGCAACAGCATGATCT | 18633 |
| Db | 1741 | AGGCAACACATCCACTCTGGGCGCCACTGCGCTTACAGAGACATGCAACAGCATGATCT | 1800 |
| QY | 1864 | GGAGATCCAAAGAGGGTAACCTGTTGGAAATCTGCGCGACGTGGGAATGGAAAAAATCTC | 19233 |
| Db | 1801 | GGAGATCCAAAGAGGGTAACCTGTTGGAAATCTGCGCGACGTGGGAATGGAAAAAATCTC | 1860 |
| QY | 1924 | TCTCATTTACGCCATTTTATAGGCGAGATGACGCTCTAGAGGGGACACATTTGCATATAGG | 19833 |
| Db | 1861 | TCTCATTTACGCCATTTTATAGGCGAGATGACGCTCTAGAGGGGACACATTTGCATATAGG | 1920 |
| QY | 1984 | AACCTTGCTTATGTGGCCCGACAGGCGCTGTGATCTCAATGTCTACTGTAGAGACACAT | 20433 |
| Db | 1921 | AACCTTGCTTATGTGGCCCGACAGGCGCTGTGATCTCAATGTCTACTGTAGAGACACAT | 1980 |
| QY | 2044 | CTGTGTTGGGAAGAGATATGATGAAGAAAGATACAACCTGTGTCTGTAACAGCTGCTGCT | 21033 |
| Db | 1981 | CTGTGTTGGGAAGAGATATGATGAAGAAAGATACAACCTGTGTCTGTAACAGCTGCTGCT | 2040 |
| QY | 2104 | GAGGCGTGACCTGGCCATTCTTCCGACAGCGCACCTGCAGGAGATTGGAGACGAGGAGC | 21633 |
| Db | 2041 | GAGGCGTGACCTGGCCATTCTTCCGACAGCGCACCTGCAGGAGATTGGAGAGGAGAGC | 2100 |
| QY | 2164 | CAACCTGAGGGTGGGCGAGCGCAGAGAGATCAACCTTGGCCCGGGCTGTATAGAGAC | 22233 |
| Db | 2101 | CAACCTGAGGGTGGGCGAGCGCAGAGAGATCAACCTTGGCCCGGGCTGTATAGAGAC | 2160 |
| QY | 2224 | GAGCATCTACATCTCTGAGGAGCAACCCCTCACTGCTTATGATGCCATGTGGGCACACAT | 22833 |
| Db | 2161 | GAGCATCTACATCTCTGAGGAGCAACCCCTCACTGCTTATGATGCCATGTGGGCACACAT | 2220 |
| QY | 2284 | CTTCAATATAGGTCATCTCCGGAAACATCCCAAGTCCAAAGAAAGTCTGTTGTTACCCACA | 23433 |
| Db | 2221 | CTTCAATATAGGTCATCTCCGGAAACATCCCAAGTCCAAAGAAAGTCTGTTGTTACCCACA | 2280 |
| QY | 2344 | GTTACAGTACCTGGTTACACTGTATGATGAAGTATCTTCATGAAGAAGAGGGCTGTATTACGA | 24033 |
| Db | 2281 | GTTACAGTACCTGGTTACACTGTATGATGAAGTATCTTCATGAAGAAGAGGGCTGTATTACGA | 2340 |
| QY | 2404 | AAGAGGCAACCCATGAGGAGTGAATTAATTAATGAGTACTTACTCATATTTTAAATAA | 24633 |
| Db | 2341 | AAGAGGCAACCCATGAGGAGTGAATTAATTAATGAGTACTTACTCATATTTTAAATAA | 2400 |
| QY | 2464 | CTGTGCTGGGAGAGACACGGCAGTGTGAGTCAATTAATAAAAGGAAACCAAGGGTTC | 25233 |
| Db | 2401 | CTGTGCTGGGAGAGACACGGCAGTGTGAGTCAATTAATAAAAGGAAACCAAGGGTTC | 2460 |
| QY | 2524 | ACAAAGAAAGTCAACAACAAGAGGTCTCTAAACAGGATCAATTAAGAAGAAAAAGCAGT | 25833 |
| Db | 2461 | ACAAAGAAAGTCAACAACAAGAGGTCTCTAAACAGGATCAATTAAGAAGAAAAAGCAGT | 2520 |
| QY | 2584 | AAAGCCAGAGAGAGGCGACTTGTGCACTGTGAGAGAAAAAGGCGTTCAGTCCCTG | 26433 |

|||||
Db 2521 AAGCCAGAGGAGGGGAGCTTGTGCAGCTGGAAAGAAAGGAGGCTAGTCCCTG 2580
OY 2644 GTCACATATGCTGCTACATCCAGGCTGCTGGGGCCCCCTTGGCATCTCTGTTATAT 2703
|||||
Db 2581 GTCACTAATATGGTGCTCTACATCCAGGCTGCTGGGGCCCCCTTGGCATCTCTGTTATAT 2640
OY 2704 GGGCTTTTCATGCTGATGATGAGGAGCAGCCCTTCAGCACCTGCTGGTTAGTACG 2763
|||||
Db 2641 GGCCCTTTTCATGCTGATGATGAGGAGCAGCCCTTCAGCACCTGCTGGTTAGTACG 2700
OY 2764 GATCAAGCAAGAGAGGAGGAGAACACCACCTGTGACCTGAGGAGAACGAGACCTCGGTAGTGA 2823
Db 2701 GATCAAGCAAGAGAGGAGGAGAACACCACCTGTGACCTGAGGAGAACGAGACCTCGGTAGTGA 2760
OY 2824 CAGCATGAAGCAATCTCTCATATGCTACTATGCTCCAGCATCTACGCCCTCTCCATGGC 2883
|||||
Db 2761 CAGCATGAAGCAATCTCTCATATGCTACTATGCTCCAGCATCTACGCCCTCTCCATGGC 2820
OY 2884 AGTCATGCTGATCCCTGAAGGCCATTCGAGGAGTTGCTTTTGTCAAGGGGAGCGCTGCGAGC 2943
|||||
Db 2821 AGTCATGCTGATCCCTGAAGGCCATTCGAGGAGTTGCTTTTGTCAAGGGGAGCGCTGCGAGC 2880
OY 2944 TTCCCTCCCGGCTGCATGACGAGCTTTTCCGAAGATCCCTTCGAAGCCCTATGAACTTTT 3003
Db 2881 TTCCCTCCCGGCTGCATGACGAGCTTTTCCGAAGATCCCTTCGAAGCCCTATGAACTTTT 2940
OY 3004 TGACACGACCCCCACAGGAGAGATTCCTCAACAGTTTTTCCAAAGACATGATGAAGTTGA 3063
Db 2941 TGACACGACCCCCACAGGAGAGATTCCTCAACAGTTTTTCCAAAGACATGATGAAGTTGA 3000
OY 3064 CGTGGGCTGGCCCTCCAGGCCGAGATGTTCTCCAGAAACCTTATCCGCTGCTCTCG 3123
Db 3001 CGTGGGCTGGCCCTCCAGGCCGAGATGTTCTCCAGAAACCTTATCCGCTGCTCTCG 3060
OY 3124 TGTGGAAATGATGCGAGAGTCTTCCCGTGTTCTTGTGGCAGTGGGGCCCCCTTGTCAAT 3183
Db 3061 TGTGGAAATGATGCGAGAGTCTTCCCGTGTTCTTGTGGCAGTGGGGCCCCCTTGTCAAT 3120
OY 3184 CCTCTTTTACGTCTGACACATTTGCTCCAGGGTCTGATTCGGGAGCTGAAGCGTCTGGA 3243
Db 3121 CCTCTTTTACGTCTGACACATTTGCTCCAGGGTCTGATTCGGGAGCTGAAGCGTCTGGA 3180
OY 3244 CAATATCAAGCGATGACCTTCTCTCCACATCACTGTCACATACAGGGGCTTGGCCAC 3303
Db 3181 CAATATCAAGCGATGACCTTCTCTCCACATCACTGTCACATACAGGGGCTTGGCCAC 3240
OY 3304 CATTCACGCTACATTAAGAGGAGAGAGTTTCTGCACAGATACAGAGACTCTGGATGA 3363
Db 3241 CATTCACGCTACATTAAGAGGAGAGAGTTTCTGCACAGATACAGAGACTCTGGATGA 3300
OY 3364 CAACCAAGCTCTTTTTTTTTTTTTTGTCTTAACTGTGCGATGCGGGTGGCTGTGCGGCTGGA 3423
Db 3301 CAACCAAGCTCTTTTTTTTTTTTTTGTCTTAACTGTGCGATGCGGGTGGCTGTGCGGCTGGA 3360
OY 3424 CCTCATACGATACGCTCATACACCAACGAGGGGATGATGCTTCTTATGCAAGGGGA 3483
Db 3361 CCTCATACGATACGCTCATACACCAACGAGGGGATGATGCTTCTTATGCAAGGGGA 3420
OY 3484 GATTCGCCAGGCTATGGGGGTCTGCGCATCTCTTATGCTGTCCAGTTAAACGGGGCTGTT 3543
Db 3421 GATTCGCCAGGCTATGGGGGTCTGCGCATCTCTTATGCTGTCCAGTTAAACGGGGCTGTT 3480
OY 3544 CCAGTTTACGCTGACAGCTGCGCATCTGAGACAGAGCTGATTCACCTCGGTGGAGAGAT 3603
Db 3481 CCAGTTTACGCTGACAGCTGCGCATCTGAGACAGAGCTGATTCACCTCGGTGGAGAGAT 3540
OY 3604 CAATACATACATTAAGAGCTGTGCTGGAAAGCAGCTGCAATTAAGAAAGAGCTCC 3663
Db 3541 CAATACATACATTAAGAGCTGTGCTGGAAAGCAGCTGCAATTAAGAAAGAGCTCC 3600
OY 3664 CTCCCTGACTGGCCCCAGAGGAGAGTGAACCTTTGAAACGACAGATGAGTTACCG 3723
|||||

Db 3601 CTCCCTGACTGGCCCCAGAGGAGAGTGAACCTTTGAGAAACGACAGATGAGTTACCG 3660
OY 3724 AGAAACCTCCCTCTGCTGCTTAAAGAAATGATCTTCACGATCAAAACCTTAAAGAGAGAT 3783
Db 3661 AGAAACCTCCCTCTGCTGCTTAAAGAAATGATCTTCACGATCAAAACCTTAAAGAGAGAT 3720
OY 3784 TGGCATTTGGGGGCGAGAGATCAAGGAAAGTCTCGCTGGGGATGAGCCCTTCCGCT 3843
Db 3721 TGGCATTTGGGGGCGAGAGATCAAGGAAAGTCTCGCTGGGGATGAGCCCTTCCGCT 3780
OY 3844 GGTGAGATTATCTGAGGCTGCATCAAGATTGATGAGTGAATGATATTTGGCCT 3903
Db 3781 GGTGAGATTATCTGAGGCTGCATCAAGATTGATGAGTGAATGATATTTGGCCT 3840
OY 3904 TGCCCACTCCCAACCAAACTCTATCATCTTCCAAAGCCGGGCTGTACGTGGGAC 3963
Db 3841 TGCCCACTCCCAACCAAACTCTATCATCTTCCAAAGCCGGGCTGTACGTGGGAC 3900
OY 3964 TGTCAATCAATTTGAGACCCCTTCACACAGTACACTGAAAGCAGATTTGGATGCCCT 4023
Db 3901 TGTCAATCAATTTGAGACCCCTTCACACAGTACACTGAAAGCAGATTTGGATGCCCT 3960
OY 4024 GGAGAGACACACATGAAGAAATGATTGCTCAGTACCTCTGAAACTTGAATCGAAGT 4083
Db 3961 GGAGAGACACACATGAAGAAATGATTGCTCAGTACCTCTGAAACTTGAATCGAAGT 4020
OY 4084 GATGAGAAATGGGGATTAATTTCTCAGTGGGGGAACGGACCTTGTGCTATGCTAGAC 4143
Db 4021 GATGAGAAATGGGGATTAATTTCTCAGTGGGGGAACGGACCTTGTGCTATGCTAGAC 4080
OY 4144 CCTGCTCCGCCACTGTAAGATTCTGATTTTATGATGAAGCACAGCTGCCATGACACAGA 4203
Db 4081 CCTGCTCCGCCACTGTAAGATTCTGATTTTATGATGAAGCACAGCTGCCATGACACAGA 4140
OY 4204 GACAGACTTATTTGATTCAGAGACCATCCGAGAGCAATTTGACAGCTGTACATGCTGAC 4263
Db 4141 GACAGACTTATTTGATTCAGAGACCATCCGAGAGCAATTTGACAGCTGTACATGCTGAC 4200
OY 4264 CATTTGCCATTCGCTTCGACACAGGTTCTAGGCTCCGATGAGATTATGCTGCTGGCCCAAGG 4323
Db 4201 CATTTGCCATTCGCTTCGACACAGGTTCTAGGCTCCGATGAGATTATGCTGCTGGCCCAAGG 4260
OY 4324 ACAGGTGGGAGTTTGAACACCCCATCGGCTCTGTGCTCAACGACAGTCCCGATTCTA 4383
Db 4261 ACAGGTGGGAGTTTGAACACCCCATCGGCTCTGTGCTCAACGACAGTCCCGATTCTA 4320
OY 4384 TGCCATGTTTGTCTGTCAGAGAACAGGCTGCTGTCAAGGGGCTGACTCTCTGTTGA 4443
Db 4321 TGCCATGTTTGTCTGTCAGAGAACAGGCTGCTGTCAAGGGGCTGACTCTCTGTTGA 4380
OY 4444 CGAAGTCTTTTCTTTTGTAGAGATTTGCCATTTCCCTGCTGGGGGCGGCTTCATCGCG 4503
Db 4381 CGAAGTCTTTTCTTTTGTAGAGATTTGCCATTTCCCTGCTGGGGGCGGCTTCATCGCG 4439
OY 4504 TCCCTCAACGAAACCTTTCCTTCGATTTTATCTTTCGACAGAGATTCGGATTGG 4563
Db 4440 TCCCTCAACGAAACCTTTCCTTCGATTTTATCTTTCGACAGAGATTCGGATTGG 4499
OY 4564 CTGTGTGTTTCACTTTTGAAGGAGATCATATTTGATTTATTTATTTTCCATATTTCA 4623
Db 4500 CTGTGTGTTTCACTTTTGAAGGAGATCATATTTGATTTATTTATTTTCCATATTTCA 4559
OY 4624 TGTAAACAAATTTTGTGTTTGTCTTAATTTGACTCTTAAAGGTTTCAGGGAACCTTAT 4683
Db 4560 TGTAAACAAATTTTGTGTTTGTCTTAATTTGACTCTTAAAGGTTTCAGGGAACCTTAT 4619
OY 4684 TATTAATTGATCAGAGGCTTATTAATGAAGCTTTATACGTTAGCTATATCTATATATAT 4743
Db 4620 TATTAATTGATCAGAGGCTTATTAATGAAGCTTTATACGTTAGCTATATCTATATATAT 4679
OY 4744 TCTGTACATAGCCCTATATTTACAGTGAAGAAATGAAGCTGTTTTATTTTATTTAAATTAAG 4803
Db 4680 TCTGTACATAGCCCTATATTTTACAGTGAAGAAATGAAGCTGTTTTATTTTATTTAAATTAAG 4739
|||||

[illegible]

| | | | | | |
|-----------------------|-----|---|-------------|----------------|-------------|
| Query Match | | 77.9% | Score 3774 | DB 94 | Length 5400 |
| Best Local Similarity | | 88.0% | Pred. No. 0 | | |
| Matches 4203 | | Conservative | 0 | Mismatches 555 | Indels 16 |
| | | | | Gaps | 8 |
| QY | 1 | GGCTCATGCTCCGGAGCGCTGTTAGAGCGCTGGCGCGGTTCTCTCGACGACGAGCGCGAG | 60 | | |
| | | | | | |
| Db | 39 | GTCCGATGTTGTCGCGACGTCGCCCGGCGCTGTTGCTCTCTGTTGGGAGGACGTGGGGTGCG | 98 | | |
| QY | 61 | GAATTCGATTCGAAACCTACACAGTCTGTGAGCCCTCGGACCTCCCATCCGACAGAAAGATGAA | 120 | | |
| | | | | | |
| Db | 99 | AATTTCTGAAGTGAAGAGAGAGTCTGCA---CCCTGGTCTCTGTGTGTGTGAAGATGAA | 155 | | |
| QY | 121 | GGATATTCGACATAGGAAAAAGATATATCATCCCGACAGTCTGGGTATAGAGTGTGAGGGA | 180 | | |
| | | | | | |
| Db | 156 | AGATATTTGACATGGGAAAAAAGATATATCATCCCGACCGCTGGGTACAGAAAGTACAGGGA | 215 | | |
| QY | 181 | GAGAACGACGACTTCTGGGACCGCACAGAGACCGTGAATTTCCAAAGTTCCAGAGAACTCG | 240 | | |
| | | | | | |
| Db | 216 | CAGAGAGCCCTGTACCAAGGGCAACACAGAGACCCGAGAGATCCCAAGTTTCCGAGAAACAG | 275 | | |
| QY | 241 | ACCGTTGGAATGCCAAGATGCTTTGGAACACAGACCCCGAGCCGAGCGGCGCTCTCTCTTGA | 300 | | |
| | | | | | |
| Db | 276 | ATTCGTGGAATGCCAAGATGCTCTCGAAGACAGACCCGAGTGAAGGGGCTTTCCCTCGGA | 335 | | |
| QY | 301 | TGCCTCCATGCATTCCTCAGCTCAGATCCTGGATGAGAGCATCCCAAGGAAAGTACCA | 360 | | |
| | | | | | |
| Db | 336 | TATCTCTGTGATTCCTCATCTCAATCCAAATTCGTGACGAGGAGATTCCTAAGGGAAAAATACCA | 395 | | |
| QY | 361 | TCAATGCGCTTGATGCTGTCAGAGCCCATCCGAGCTCTTGCACAAACACAGACCCAGTGA | 420 | | |
| | | | | | |
| Db | 396 | CCATGCTTAAAGTGTCTCGAAGACCTTCCGGACACACTACAAAGACACAGACCCAGTGA | 455 | | |
| QY | 421 | CAATGCTGGGCTTTTTCCTGTATGACTTTTTCGTGCTTCTCTCTGCGCCCGTGTGTC | 480 | | |
| | | | | | |
| Db | 456 | CAATGCTGACTTTTCTCTCTCACTACACTTTTTCATGAGCTCTCTCTGCGCCGAGTGT | 515 | | |
| QY | 481 | CCACAAAGAGGGGAGCTCTCAATGGAAGAGCTGTGTCTGTCTCAAGACAGACTCTTC | 540 | | |
| | | | | | |
| Db | 516 | TCACAAAGAGGGGAGCTGTATATGGAAGATGTGTGCGCTTTGTCCAAAGATGAATCTTC | 575 | | |
| QY | 541 | TGACGTGAACAGTCAGAAAGACTAGAGAGACTGTGCGCAAGAGACTGAATGAATGGGACC | 600 | | |
| | | | | | |
| Db | 576 | TGATGTGAACAGTCAGAAAGACTAGAGAGACTGTGTGCGCAAGAGCTGAATGAATGGGACC | 635 | | |
| QY | 601 | AGAGCTGCTCTCTCTGCGCAAGGCTGTGTGATCTTCTGTGCGCACACAGGCTCATCTGTCT | 660 | | |
| | | | | | |
| Db | 636 | AGAGCTGCTCTCTCTGCGCAAGGCTGTGTGATCTTCTGTGCGCACACAGGCTCATCTGTCT | 695 | | |
| QY | 661 | CATGCTGTGCGCTGATGATCAGCAGCAGCTGTGTGCGCTCAAGTGAGACAGGCTTCATGTGTA | 720 | | |
| | | | | | |
| Db | 696 | CATGCTGTGCGCTGATGATCAGCAGCAGCTGTGTGCGCTCAAGTGAGACAGGCTTCATGTGTA | 755 | | |
| QY | 721 | ACACCTCTTGGAGTATACCCAGGCAACAGAGTCTAACCCTGACAGTACAGCTTGTGTAGT | 780 | | |
| | | | | | |
| Db | 756 | ACACCTCTTGGAGTATACCCAGGCAACAGAGTCTAACCCTGACAGTACAGCTTGTGTAGT | 815 | | |
| QY | 781 | GCTGGGCTCTCTCTGAGAGGAAATGTGTGCGTCTGTGTGCTTGTGACATGACTTGGGGCAAT | 840 | | |
| | | | | | |
| Db | 816 | GCTGGGCTCTCTCTGAGAGGAAATGTGTGCGTCTGTGTGCTTGTGACATGACTTGGGGCAAT | 875 | | |
| QY | 841 | GAATTAACGGAACCGGTGTGTCGCTTCGGGGGGCCATCTCTAACCATAGGGAATTTAAAGAAAT | 900 | | |
| | | | | | |
| Db | 876 | GAATTAACGGAACCGGTGTGTCGCTTCGGGGGGCCATCTCTAACCATAGGGAATTTAAAGAAAT | 935 | | |
| QY | 901 | CCTTAATGTTAAAGAACTTAAAGAGAAATCCCTGGGTGAAGTCAATCAATTTGCTCCAA | 960 | | |
| | | | | | |
| Db | 936 | CCTTAATGTTAAAGAACTTAAAGAGAAATCCCTGGGTGAAGTCAATCAATTTGCTCCAA | 995 | | |
| QY | 961 | CGATGGGAGAGAAATGTTGAGGCGACAGCGCTTGGACAGCGCTGTGGCGTGAAGACCCGT | 1020 | | |
| | | | | | |
| Db | 996 | CGATGGGAGAGAAATGTTGAGGCGACAGCGCTTGGACAGCGCTGTGGCGTGAAGACCCGT | 1055 | | |

| | | | | | |
|----|------|--------------|------------------------|---------------------------------------|------|
| QY | 1021 | TGTTGCCATCTT | TAGCATGATTTATATATGTAA | TATCTGGAGCCACACAGCCTTCTG6G | 1080 |
| Db | 1056 | TGTTGCCATCTT | TAGCATGATTTATATATGTAA | TATCTGGAGCCACACAGCCTTCTG6G | 1115 |
| QY | 1081 | ATCAGCTGTTTT | TATCCTCTTTTACCACACAT | TGATGTGTTTGATACGGCTCACAGCAT | 1140 |
| Db | 1116 | ATCAGCGTTTTT | TATCCTCTTTTATTCACACAT | TGATGTGTTTGATACGGCTCACAGCAT | 1175 |
| QY | 1141 | TTTTAGAGAAAA | TGCGTGGCCGCCACGAGTAA | AGTGTCGCGAAGATGAAGTGTCT | 1200 |
| Db | 1176 | TTTTAGAGAAAA | TGCGTGTAGCTGTAGCTGCCAC | AGATATACCGTGTGCCAAGATGAAGTGTCT | 1235 |
| QY | 1201 | TACTTACATTAA | TTTTATCAAAATGTATGSC | TGGGTCAAAGCATTTTCTCAGAGTGTTCA | 1260 |
| Db | 1236 | TACCTACATTAA | TTTTATCAAAATGTATGATC | TGGGTCAAAGCGTTTCTCAGTGTGTGCA | 1295 |
| QY | 1261 | GAAATCCCGAG | AGAGAGAGCGTGGATATTGG | AAAAAGCCGGGTACTTCCAGAGCATCAC | 1320 |
| Db | 1296 | AAAAATCCGAG | AGAGAGAGAGTGGATTTGG | AAAAAGCCGGGTACTTCCAGAGCATCAC | 1355 |
| QY | 1321 | TGTGGGTGTGG | CTCCCATTTGCGTGTGATATTC | CGACGCGTGAGCACTTCTCGTTCAATAT | 1380 |
| Db | 1356 | TGTTGGAGTGT | GGCTCTATTTGTGTGTGTGTGT | GTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1415 |
| QY | 1381 | GACCCCTGGG | CTTGATCTGCACAGCAGCAC | AGGCTTTCACAGTGTGTGACAGTCTTCAATTC | 1440 |
| Db | 1416 | GACCCCTGGG | CTTGATCTGCATCTGCCTG | CGGCACAGGCGTTTCAAGTGTGTGTGTGTGTGT | 1475 |
| QY | 1441 | CATGACTTTT | GCTTTGAAAGTAAACACCC | GTTTTCAGTAAAGTCCCTCTCAGAAAGCTCTCAGT | 1500 |
| Db | 1476 | CATGACTTTT | GCTTTGAAAGTAAACACAT | TTCAGTAAAGTCCCTCTCTCAGAAAGCTCTCAGT | 1535 |
| QY | 1501 | GGCGTTTGAC | AGATTTAAGGTTTGTTTCTA | ATNGAAGAGGTTCACATGATTAAGAAACAA | 1560 |
| Db | 1536 | GGCGTTTGAC | AGATTTAAGGTTTGTTTCTA | ATNGAAGAGGTTCACATGATTAAGAAACAA | 1595 |
| QY | 1561 | ACGACGCA | GTCTCACATCAAGATTAAG | ATGAAGTGCACCTTGGGATGGGACCTCTC | 1620 |
| Db | 1596 | ACGCGGCA | GTCTCACATCAAGATTAAG | ATGAAGTGCACCTTGGGATGGGACCTCTC | 1655 |
| QY | 1621 | CCACTCCAG | TATCCAGAACTGCGCC | CAAGCTGACCCCCCAAAATGAAAAAAGACAGAGGGC | 1680 |
| Db | 1656 | CCACTCCAG | TATCCAGAACTGCGCC | CAAGCTGACCCCCCAAAATGAAAAAAGACAGAGGGC | 1715 |
| QY | 1681 | TTTCAGAGGG | CAAAAGAGAGGTGAGG | AGCAGTGCAGCGCATCAGAGGGGTCT | 1740 |
| Db | 1716 | TTCAGAGGG | CAAAAGAGAGGTGAGG | AGCAGTGCAGCGCATCAGAGGGGTCT | 1775 |
| QY | 1741 | GGCAGACAG | AAAAAGGCCACTCTCT | CTGTGACAGTGCAGCGGGCCAGTCCCGAAGAGGA | 1800 |
| Db | 1776 | GGCAGACAG | AAAAAGGACACTCTCTCT | GTGACAGTGCAGCGGGCCAGTCCCGAAGAGGA | 1835 |
| QY | 1801 | AGAGAGCA | GCACATCCACTTGG | CCACTCTTGACAGAGACACTGCACAGCATGA | 1860 |
| Db | 1836 | AGAGAGCA | GCACATCCACTTGG | CCACTCTTGACAGAGACACTGCACAGCATGA | 1895 |
| QY | 1861 | TTCTGGAAT | CCAAAGAGGTAACTG | TGTTGGAATCTGCGCAGTGTGGGAAGTGA | 1920 |
| Db | 1896 | CTTGAATTA | TTGAAGAGGCAAACT | GTGTTGGAATCTGCGCAGTGTGGGAAGTGA | 1955 |
| QY | 1921 | CTCCTCAT | TTTACGCAATTTAG | CCAGATGACGCTTCTAGAGGGCAGCATTTGCAATAG | 1980 |
| Db | 1956 | CTCCTCAT | TTTACGCAATTTAG | CCAGATGACGCTTCTTGGAGGGCAGCATTTGCGGTAG | 2015 |
| QY | 1981 | TGGAACCT | TGCTTATGTGCGCC | CAGAGCCTGAGATCTCAATGCTACTGTAGAGACAA | 2040 |
| Db | 2016 | TGGAGCTT | TGCTTATGTGCGCC | CAGAGCCTGAGATCTCAATGCTACTGTAGAGACAA | 2075 |
| QY | 2041 | CATCCTGT | TTTGGGAAGATATGAT | GAAGAAATCAACTCTGTGCTGACACAGCTGCTG | 2100 |
| Db | 2076 | CATCCTGT | TTTGGGAAGATATGAT | GAAGAAATCAACTCTGTGCTGACACAGCTGCTG | 2135 |

| | | | |
|----|------|--|------|
| QY | 2101 | CTGAGGCTCTGACCTGGCCATTCTCTCCAGACAGCCACTGACGGAGATTGAGAGCCGAG | 2160 |
| Db | 2136 | CCCTGCGGCTGTGACTTGTGGCCATTCTCCCAACAGCCAGCTGACTGGAGATTGGAGCCGAG | 2195 |
| QY | 2161 | AGCCAACTGAGGGGTGGGAGCGGACAGAGGATACACCTTGGCCGGGCTGTATAGGA | 2220 |
| Db | 2196 | AGCCAACTGAGGTGTGGACAGCGCCAGAGATCACCTTGGCTTAGACCTTGTACAGTGA | 2255 |
| QY | 2221 | CAGGAGATCTACATCCTGGAGCAGCCCTCTAGTGGCTTGAATGGCCATGTGGCCAA | 2280 |
| Db | 2256 | TAGAAAGATCTACATCCTGGATATAGCCCTCAGTGGCTTGAATGGCCATGTGGCCAA | 2315 |
| QY | 2281 | CATCTTCATATGTCTATCCGGAACATCTCAAGTCCAAGACGTTCTGTTGTTACCA | 2340 |
| Db | 2316 | CATCTTCACAGTGTCTATCCGGAAGGCTCCAAGTCTTAAGAGGTTCTGTTGTTACCA | 2375 |
| QY | 2341 | CCAGTTACAGTACCTGGTTATGAGTGTATGAAGATCTTCATGAAGAGGGCTGTATAC | 2400 |
| Db | 2376 | CCAGTTACAGTATCTGGTTCATGTGTGTGAGGTGATCTTCATGAAGAGGCTGTATAC | 2435 |
| QY | 2401 | GGAAGAGGCGCCCATATGAGACTGTATGAATTTAAATGGTGACTATGCTACCATTTTAA | 2460 |
| Db | 2436 | AGAGAGAGTACCCATGAGGAGGTGATGAACCTAAATGGGGATTACGCTACGATTTTAA | 2495 |
| QY | 2461 | TACCTGTGCTGGGAGACACCCGCACTTGATGATCAATTTAAAAAGAAACACATGG | 2520 |
| Db | 2496 | TACCTGTGCTGGGAGAGACCCCGCACTTGATGATTAATTTGAAAAAGAAAGCTAC | 2555 |
| QY | 2521 | TTTCACAGAAAGTCAACAAACAAAGGGTCTTAAACAGATCAATTAAGAAAGAAAAAGC | 2580 |
| Db | 2556 | TTTCAC--AAAAATCAACAAACAAAGGGCTTAAAGCCAGGTCAGTGAAGAAAGAAAGC | 2612 |
| QY | 2581 | AGTAAAGCCAGAGGAAAGGAGCGACTTGTGCAGCTGGAAGAAAGGCGAGTTCAGTCC | 2640 |
| Db | 2613 | GGTGAAGTCGAGAAAGGAGCGACGCTGTGCAAGTGTGAGAGAAAGGCGAGTTCAGTCC | 2672 |
| QY | 2641 | CTGCTCAGTATATGTGTCTACATCCAGGCTGCTGGGGCCCTTGCGATTCCTGTTAT | 2700 |
| Db | 2673 | TTGGCAGTCTACAGGGCTTACATCCAGGCTGCGAGGGGCCCTTGCTGCTGCTAT | 2732 |
| QY | 2701 | TATGGCCCTTTTCATGCTGTAATGTAGGACACACCGGCTTCACACCTGGTGGTATGA | 2766 |
| Db | 2733 | CATGGCTCTTTCATGCTGTAATGTGGGACACACCTGCTTCACACCTGGTGGTATGA | 2792 |
| QY | 2761 | CTGGATTCAGGAAGGAGCGGAAACACATGTGATCTCGAGGAAAGACACTCGGTAG | 2820 |
| Db | 2793 | CTGGATTCAGGAAGGAGCGGAAACACAGTGTATCAAGGGAAACAGACTTGTGTAG | 2855 |
| QY | 2821 | TGACAGCATGGAAGACATCTCATATGACAGTATGTCAGACATCTACGCCCTCTCAT | 2880 |
| Db | 2853 | TGACAGCATGGAAGACAAACCCCTTCATGACAGTACTACGCGAGATCTACGCCCTCTCAT | 2912 |
| QY | 2881 | GGCAGTCACTGCTATCTGAAAGCCATTCGAGAGTTGTCTTGTCAAGGGACGCTGCG | 2940 |
| Db | 2913 | GGCAGTCACTGCTATCTTGAAGCCATTCGAGAGTTGTCTTGTCAAGGGACGCACTAG | 2972 |
| QY | 2941 | AGCTTCCTCCCGGCTGTGATGACGAGGCTTTCCGAAGATCTCTCGAAGCCATTAAGT | 3000 |
| Db | 2973 | AGCTTCCTCCCGGCTCATATGTGAGTATCTCGAAGAGATCTTAGAGCCCATTAAGT | 3033 |
| QY | 3001 | TTTTGACAGACCCCAACAGGAGGATTTCAACAGGTTTTCCAAAGACATGGATGAAGT | 3066 |
| Db | 3033 | TTTTGATATCTACCCCAACAGGAGGATTTCAACAGGTTTTCCAAAGACATGGATGAAGT | 3092 |
| QY | 3061 | TGACGTGCGGCTGGCCTTTCCAGGCGGAGATGTTCAATCAAGAACGTTATCTGTGTTCT | 3120 |
| Db | 3093 | GGATGTGCGGCTGGCCTTTCCAGGCTGAGATGTTATTCAGAAATGTAACTGTGTTCTT | 3152 |
| QY | 3121 | CTGTGTGGGATGATGCGAGGAGCTTCCCGGTGTTCTTGTGGCAGAGTGGGCGCCCTGT | 3180 |
| Db | 3153 | CTGTGTGGGATGATGCTGGAGTCTTCCCATGTTCTCTGTGGCGGTGGGCTCTCTCT | 3212 |
| QY | 3181 | CATCTCTTTTCATCTCCGACATATGTCTCCAGGCTCGATTCGGAGCTTAAGCGTCT | 3240 |

| | | | |
|----|------|---|------|
| Db | 3213 | CATCCTCTTCCACCTTCTCCACATTTGCTCCAGGGGCTGATTTCTGGAGCTAAAGCGGTT | 3272 |
| Qy | 3241 | GGACAATATACGCGACGTCACCTTCTCTCCACATCAGTCCACCATACAGGGCTTGC | 3300 |
| Db | 3273 | GGACAATATACGCGACGTCCTTCTCTCCACATCAGTCTACCATTCAGGGGCTTGC | 3332 |
| Qy | 3301 | CACCATCAGCGCTACAAATTAAGGGCAGAGATTTCGACAGATACCAAGAGGCTCTGA | 3360 |
| Db | 3333 | CACCATCAGTCCGACAAACAAAAGGACAGGATTTTACACAGATATCAGAGGCTCTGA | 3392 |
| Qy | 3361 | TGACAACCAAGCTCTCTTTTGTTTTACGTGTGCGATGCGGTGGCTGTGCGGCT | 3420 |
| Db | 3393 | TGACAACCAAGCTCTCTTTTGTTTTACGTGTGCGATGCGGTGGCTGTGCGGCT | 3452 |
| Qy | 3421 | GGACCTCATGAGATCGCCCTTCATCCACCAAGGGGCTGATGATCGTTCTTATGCAGG | 3480 |
| Db | 3453 | GGACATCATGAGATCGCTTGTGATTTACAGACATGCGCTGATGATTTGTCTCAGCATGCG | 3512 |
| Qy | 3481 | GCAGATTCGCCCAAGCTTATCGGGGTGCGCATCTCTTATGCTGTCCAGTTAACGGGCT | 3540 |
| Db | 3513 | CAGATTCCTTTCAGGCTATCGGGGCTTGCATTTCCATCGCTGTGCAGTTAATGAGCT | 3572 |
| Qy | 3541 | GTTCACATTTACGGTTCAGACTGTGCGATCTAGACAGAGTCGATACCTCGGTGGAGAG | 3600 |
| Db | 3573 | ATTCACATTTACGGTTCAGACTGTGCGATCTAGACAGAGTCGATACCTCGGTGGAGAG | 3632 |
| Qy | 3601 | GATCAATACATTAATTAAGACTGTCTCTTGGAAAGCAGCTGCGCAATAATTAAGAACAGC | 3660 |
| Db | 3633 | GATCAACACATTAATTAAGACTGTCTCTTGGAAAGCAGCTGCGCAATAATTAAGAACAGC | 3692 |
| Qy | 3661 | TCCCTCCCTGACTGGGCCCGAGAGGAGAGGTGACTTTTGAAACCGAGAGATGAGTA | 3720 |
| Db | 3693 | TCCCTCCCTGACTGGGCCCGAGAGGAGAGGTGACTTTTGAAACCGAGAGATGAGTA | 3752 |
| Qy | 3721 | CCGAGAAAACCTCCCTCTCGCTCTTAAGAAAGATCCTTACAGATCAAAACCTTAAGAAAG | 3780 |
| Db | 3753 | CCGAGAAAACCTCCCTCTCGCTCTTAAGAAAGATCCTTACAGATCAAAACCTTAAGAAAG | 3812 |
| Qy | 3781 | GATTGGCAATGTGGGGCGAGCAGAGATCAGGGAAGTCTCTGCTGGGGATGCGCCCTTCCG | 3840 |
| Db | 3813 | GATTGGCAATGTGGGGCGAGCAGAGATCAGGGAAGTCTCTGCTGGGGATGCGCCCTTCCG | 3872 |
| Qy | 3841 | TCTGCTGAGATTATCTGAGAGCTGTCATCAAGATGTGTGAGTGAAGATACGATTAATGG | 3900 |
| Db | 3873 | TCTGCTGAGATTATCTGAGAGCTGTCATCAAGATGTGTGAGTGAAGATACGATTAATGG | 3932 |
| Qy | 3901 | CTTTGGCGACCTCCGAAACAACTCTTATCATTTCTCTCAAGAGCGGTGCTGTCAATGG | 3960 |
| Db | 3933 | CTTTGGCGACCTCCGAAACAACTGCTCATTTCTCTCAGAGGCCAGTGTCTGTCAATGG | 3992 |
| Qy | 3961 | CACGTGAGATCAAAATTTGGAACCCCTTCAACCAAGTACACTGAAAGACAGATTTGGGATGC | 4020 |
| Db | 3993 | CACGTGAGATCAAAACCTGGAACCTTTCACCAAGTACAGGAAAGCCAGATCTGGGATGC | 4052 |
| Qy | 4021 | CCGTGAGAGACACACTGAAAGAAATGTTTGTCTCAGTACTCTGAAACTGTGATCTGA | 4080 |
| Db | 4053 | TCTGAGAGAAACGCAACTGAAGAAATGTTTGTCCACAGTACTCTGAAACTGTGATCTGA | 4112 |
| Qy | 4081 | AGTGATGGAGATGGGAGTAACTTCTCACTAGTGGGGGAACGGCAGCTTGTGTGATAGTAG | 4140 |
| Db | 4113 | AGTGATGGAGAAAGGGGCAAACTTCTCTGTGTGGGGAACGGCAGCTTGTGTGATAGTAG | 4172 |
| Qy | 4141 | AGCCCTCTCCGCACTGTAAAGTATCTGATTTAGTGAAGCCACAGCTCGCATGAGAGAC | 4200 |
| Db | 4173 | AGCCCTCTCTGCACTGTAAAGTATCTGATTTAGTGAAGCTACACCGCGCATGAGAGAC | 4232 |
| Qy | 4201 | AGACACAGACTTATGATTTCAACAGACCAATCCGAGAAAGATTTCAGACTGTACCATGCT | 4260 |
| Db | 4233 | AGACACAGACTTATGATTTCAACAGACCAATCCGAGAAAGATTTCGGGAGCTGTGACCATGCT | 4292 |
| Qy | 4261 | GAGCATTCGCATGCGCTGCACACGGTCTTAGGCTCCGATAGGATTAATGTGTGCGCCA | 4320 |

[illegible]


```
homologue"
/codon_start=1
/protein_id="BAA22887.1"
/db_xref="GI:2534610"
/translation="MKNAITLAMPSSSHSSSIONSPKLTTPKMKDKKRNKSKKEKRYOLR
TERQAVLAIEQKHLIDSDERPEEKEKHILGHILRLORTLSIDLEIOEGVLGI
CGSVGSKTSLISILIGOMTLLEGISINISGFPAVAQAOAMLANLTADNLIFGEYE
ERYNSVNSGCLIPDLILPSDLEIPEGANSGORISLARLALYDRIYITID
DPLSALDAHVNHTFNSAIRKHLKSTVLTLYTHOLOYLVDDETEPKKEGCTIRGTH
EELMNLNGDYATIFENLLIGEPVEINSKETSGSKSQDKPKTSVKEKAVAP
EEQOLVLEEGGGSVPWSYGYVIOAGDPLEFVLMALFMLNVGSTASTWMLSYM
IKGGSGNTVTYRGNETSVSDSMKNPMQYASISYALSMALMLIKIRGVKLVYL
RASRLHDELFRILRSPMKPDPTPTGRILNRSKMDVDVRLPROAEFIONVL
VPCVGMIAGVPEMPELVAVGPVILVPSVLIVSVRLIRKLKRNITQSPFLSTISS
IOGLATHATNKGOEFLHRYOELDDNOAFPEFTCAKRLNLAVDLIDISTALITTEL
MYLMHGOIPPAITAGLAISTAVOLITGLFTVRLASBETAELFVSERINMHTLSLE
APARIKRNKASPMPPOGEVTFENAEKMYRENLPLVAKVSTIPEKIKIGIVRTGS
GKSISLMALRLVELSGGCIKIDVIRISDGLADLRKSLIIPQEPVLFSGTVSNID
PFIQYTEDOIMDALEERTHMEKCIQOLPLKLESEYEMNGDNFVSEROLICARLLRH
CKLILDEATAAMDTEFDLILQETIREAPADCTMLNLAHRLHTVLSDRIMVLAQGV
VERDTPSVLLSNSRFPYAFMAAENKVAANKG"
polya_site
4939
/note="15 A nucleotides"
BASE COUNT      1174 a      1224 c      1249 g      1292 t
ORIGIN
Query Match      76.4% Score 3703.8; DB 85; Length 4939;
Best Local Similarity 96.9% Pred. No. 0;
Matches 3844; Conservative 0; Mismatches 7; Indels 116; Gaps 2;

QY 968 CAGGAATGTTGAGGCGAGCGCTGGCAGCCGTCGGCTGGAGGAGCCGCTTGGCC 1027
    |||||||
Db 1 CAGGAATGTTGAGGCGAGCGCTGGCAGCCGTCGGCTGGAGGAGCCGCTTGGCC 60

QY 1028 ATCTAGGCATGATTTAATATGTAATATCTTGGGACCAAGAGCGCTTCGGATCAGCT 1087
    |||||||
Db 61 ATCTAGGCATGATTTAATATGTAATATCTTGGGACCAAGAGCGCTTCGGATCAGCT 120

QY 1088 GTTTTATACCTTTTACCAGCAATGATTTTGCATCAGCGCTCAGCAATATTTCAAG 1147
    |||||||
Db 121 GTTTTATACCTTTTACCAGCAATGATTTTGCATCAGCGCTCAGCAATATTTCAAG 180

QY 1148 AGAAAATCGTGGCGCCGACGAGATGATGCTCCAGAGATGAATGAATGTTCTTCTTAC 1207
    |||||||
Db 181 AGAAAATCGTGGCGCCGACGAGATGATGCTCCAGAGATGAATGAATGTTCTTCTTAC 240

QY 1208 ATTTAAATTTATCAAAATGATGCTGGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAAATC 1267
    |||||||
Db 241 ATTTAAATTTATCAAAATGATGCTGGGTCAAAAGCATTTTCTCAGAGTGTTCAGAAAATC 300

QY 1268 CGCGAGGAGGAGCGTCCGATATTTGAAAAAGCCGGTACTTCCAGAGATCAGTCTGGGT 1327
    |||||||
Db 301 CGCGAGGAGGAGCGTCCGATATTTGAAAAAGCCGGTACTTCCAGAGATCAGTCTGGGT 360

QY 1328 GTGGCTCCCATTTGTTGGTGTATGTCGACGCGTGGACCTTCTGTCATATGACCCG 1387
    |||||||
Db 361 GTGGCTCCCATTTGTTGGTGTATGTCGACGCGTGGACCTTCTGTCATATGACCCG 420

QY 1388 GGCCTTCGATCTGACAGCAGCAGAGCTTTCACAGTGTGACAGTCTTCAATTCATGACT 1447
    |||||||
Db 421 GGCCTTCGATCTGACAGCAGCAGAGCTTTCACAGTGTGACAGTCTTCAATTCATGACT 480

QY 1448 TTTGCTTTGAAAGTAAACCGCTTTTCACTAAAGTCCCTCTCAGAAAGCCTCAGTGGCTGT 1507
    |||||||
Db 481 TTTGCTTTGAAAGTAAACCGCTTTTCACTAAAGTCCCTCTCAGAAAGCCTCAGTGGCTGT 540

QY 1508 GACGAGATTTA----- 1517
    |||||||
Db 541 GACGAGATTTAAGCTTCTTCCACCTGTATAGCCTGATGTGTTTGTCTACGTAGTGTACGC 600

QY 1518 ----- 1517
```

```
Db 601 CCTAGGCTTGTCTCCGCTGTGTATCCAGTGTCTCATCTTTGACCTCTCTCCACAACTC 660

QY 1518 -----AGAGTTTGTTCATATGGAAGAGGCTTCACATGTATTAAGAACAAACCAAGCTCC 1572
    |||||||
Db 661 CTATAGAGTTTGTTCATATGGAAGAGGCTTCACATGTATTAAGAACAAACCAAGCTCC 720

QY 1573 TCACATCAAGATAGAGATGAAGAAAATGCCACCTTGCGATGGGACCTCTCCACATCAGTAT 1632
    |||||||
Db 721 TCACATCAAGATAGAGATGAAGAAAATGCCACCTTGCGATGGGACCTCTCCACATCAGTAT 780

QY 1633 CCAGAACTCGGCCCAAGCTGACCCCAATGAAAAAAGACAAGAGGCGCTTCAGGGGCAA 1692
    |||||||
Db 781 CCAGAACTCGGCCCAAGCTGACCCCAATGAAAAAAGACAAGAGGCGCTTCAGGGGCAA 840

QY 1693 GAAAGAGAAGTGAAGGCACTGACGCACTGAGCATATAGCGGCTGCTGGCAGAGCAAA 1752
    |||||||
Db 841 GAAAGAGAAGTGAAGGCACTGACGCACTGAGCATATAGCGGCTGCTGGCAGAGCAAA 900

QY 1753 AGGCCACCTCCTCGGACAGTGAAGAGAGCGGCGCCAGTCCCGAAGAGAAAGCAAGCA 1812
    |||||||
Db 901 AGGCCACCTCCTCGGACAGTGAAGAGAGCGGCGCCAGTCCCGAAGAGAAAGCAAGCA 960

QY 1813 CATCCACCTGGGCGCACTGCGCTTACAGAGGACACTGCACAGCATGATCTGGAGATCCA 1872
    |||||||
Db 961 CATCCACCTGGGCGCACTGCGCTTACAGAGGACACTGCACAGCATGATCTGGAGATCCA 1020

QY 1873 AGAGGTTAACTGGTTGGAATCTGCGGCACTGTGGAAAGTGGAAAAACCTCTCTATTTTC 1932
    |||||||
Db 1021 AGAGGTTAACTGGTTGGAATCTGCGGCACTGTGGAAAGTGGAAAAACCTCTCTATTTTC 1080

QY 1933 AGCCATTTTAGGCGCAGATGACGCTCTGAGAGGCGCAATGCAATGCAATGGAACCTTGGC 1992
    |||||||
Db 1081 AGCCATTTTAGGCGCAGATGACGCTCTGAGAGGCGCAATGCAATGCAATGGAACCTTGGC 1140

QY 1993 TTATGTGGCCCAAGAGCGCTGTGATCTCAATGTCTACTGTAGAGACAACATCTGTTGG 2052
    |||||||
Db 1141 TTATGTGGCCCAAGAGCGCTGTGATCTCAATGTCTACTGTAGAGACAACATCTGTTGG 1200

QY 2053 GAAGAAATATGATGAAGAAAGATTAACCTGTGTGAAACGCTCTCGTAGGCGCTGGA 2112
    |||||||
Db 1201 GAAGAAATATGATGAAGAAAGATTAACCTGTGTGAAACGCTCTCGTAGGCGCTGGA 1260

QY 2113 CCTGSCCATTTCTCCACAGAGGACCTGACGAGATTTGAGAGAGGAGGCAACCTGAG 2172
    |||||||
Db 1261 CCTGSCCATTTCTCCACAGAGGACCTGACGAGATTTGAGAGAGGAGGCAACCTGAG 1320

QY 2173 CGGTGGGAGCGCCAGAGATCAGGCTTGGCCCGGCGCTTGTAATGATGACAGAGCATCTA 2232
    |||||||
Db 1321 CGGTGGGAGCGCCAGAGATCAGGCTTGGCCCGGCGCTTGTAATGATGACAGAGCATCTA 1380

QY 2233 CATCTGAGACGACCCCTCAGTGTCTTATGATGCCCATGTGGGCAACCATCTTCAATAG 2292
    |||||||
Db 1381 CATCTGAGACGACCCCTCAGTGTCTTATGATGCCCATGTGGGCAACCATCTTCAATAG 1440

QY 2293 TGCTATCCGGAAACATCTCAAGTCCCAAGACATGTTGTTTGTACCCACATTTACAGTA 2352
    |||||||
Db 1441 TGCTATCCGGAAACATCTCAAGTCCCAAGACATGTTGTTTGTACCCACATTTACAGTA 1500

QY 2353 CCTGTTGACTGTGATGAAGTATCTTCAATGAAGAGGCGCTGATTTACGAAAGAGGCAC 2412
    |||||||
Db 1501 CCTGTTGACTGTGATGAAGTATCTTCAATGAAGAGGCGCTGATTTACGAAAGAGGCAC 1560

QY 2413 CCATAGAGAACTGATGAATTTAATGTGACTATGCTACATTTTAAATTAACCTTGTCT 2472
    |||||||
Db 1561 CCATAGAGAACTGATGAATTTAATGTGACTATGCTACATTTTAAATTAACCTTGTCT 1620

QY 2473 GGGAGAGACACCGGAGTGTGATCAATTTCAAAAAAGAAACAGTGGTGCACAGAA 2532
    |||||||
Db 1621 GGGAGAGACACCGGAGTGTGATCAATTTCAAAAAAGAAACAGTGGTGCACAGAA 1680

QY 2533 GTACAAAGACAAAGGCTCTTAAACAGAGATCAATTAAGAAAGAAAGCAAGTAAAGCCAGA 2592
    |||||||
Db 1681 GTACAAAGACAAAGGCTCTTAAACAGAGATCAATTAAGAAAGAAAGCAAGTAAAGCCAGA 1740
```

| | | | |
|----|------|---|------|
| QY | 2593 | GGAGGGGACACTTGTGCAGCTGGGAAGAAGAAAGGCGAGGGTTCCAGTGGCCCTGGTCAGATATA | 2652 |
| Db | 1741 | GGAAAGGGGACACTTGTGCAGCTGGGAAGAAGAAAGGCGAGGGTTCCAGTGGCCCTGGTCAGATATA | 1800 |
| QY | 2653 | TGAGTGTACATCCAGGCTCTGGGGGCCCTTGGCCATTCCTGGTATTATTAAGGCCCTTTT | 2712 |
| Db | 1801 | TGAGTGTACATCCAGGCTCTGGGGGCCCTTGGCCATTCCTGGTATTATTAAGGCCCTTTT | 1860 |
| QY | 2713 | CATGCTGAATGTAGCGAGCACCGCCCTTCAGCACCTGTGTGTGAATTACTGGATTCAGACA | 2772 |
| Db | 1861 | CATGCTGAATGTAGCGAGCACCGCCCTTCAGCACCTGTGTGTGAATTACTGGATTCAGACA | 1920 |
| QY | 2773 | AGGAAGGGGGAACACACACCTGTGACTGGAGGAAACAGACACTCTGGTGAAGTACAGATGAA | 2832 |
| Db | 1921 | AGGAAGGGGGAACACACCTGTGACTGGAGGAAACAGACACTCTGGTGAAGTACAGATGAA | 1980 |
| QY | 2833 | GGAAATATCTCATATGCGAGTACTATCCAGACATCTACGCCCTCTCCATGAGCGATCATGCT | 2892 |
| Db | 1981 | GGAAATATCTCATATGCGAGTACTATCCAGACATCTACGCCCTCTCCATGAGCGATCATGCT | 2040 |
| QY | 2893 | GATCTTGAAAGCCATTCGAGAGTGTCTTTTGTCAAAGGCGACGCTGCGAGCTTCCTCCCG | 2952 |
| Db | 2041 | GATCTTGAAAGCCATTCGAGAGTGTCTTTTGTCAAAGGCGACGCTGCGAGCTTCCTCCCG | 2100 |
| QY | 2953 | GCCTCATAGACAGCTTTTCCGAGGATTCCTTGGAAAGCCCTATGAGATTTTTCACAGAC | 3012 |
| Db | 2101 | GCCTCATAGACAGCTTTTCCGAGGATTCCTTGGAAAGCCCTATGAGATTTTTCACAGAC | 2160 |
| QY | 3013 | CCCCACAGGGAGGATTTCTCAACAGGTTTTCCAAAGACATGATGTAAGTTGACGTCGGCT | 3072 |
| Db | 2161 | CCCCACAGGGAGGATTTCTCAACAGGTTTTCCAAAGACATGATGTAAGTTGACGTCGGCT | 2220 |
| QY | 3073 | GCCGTTCCAGCGCGAGATGTTCAATCCAGAAGCTATCTGTGTCTTCTGTGTGGAAAT | 3132 |
| Db | 2221 | GCCGTTCCAGCGCGAGATGTTCAATCCAGAAGCTATCTGTGTCTTCTGTGTGGAAAT | 2280 |
| QY | 3133 | GATGCGAGGAGTCTCCCGGATGTTCTTGTGGCAGAGGGGGCCCTGTGCATCCCTTTTC | 3192 |
| Db | 2281 | GATGCGAGGAGTCTCCCGGATGTTCTTGTGGCAGAGGGGGCCCTGTGTGCATCCCTTTTC | 2340 |
| QY | 3193 | AGTCCCTGCACATTTGTCTCCAGGGTCTCTGATTTGGGAGCTGAAGGCTCTGCGACAATATCAC | 3252 |
| Db | 2341 | AGTCCCTGCACATTTGTCTCCAGGGTCTCTGATTTGGGAGCTGAAGGCTCTGCGACAATATCAC | 2400 |
| QY | 3253 | GCAATCACCTTTCTCTCCCAKATCACAGCTCCAGACATACAGGGGCTTGGCACCATCCACGC | 3312 |
| Db | 2401 | GCAATCACCTTTCTCTCCCAKATCACAGCTCCAGACATACAGGGGCTTGGCACCATCCACGC | 2460 |
| QY | 3313 | CTACAAATAAAGGCGAGGATTTCTTGACAGATACACAGAGGCTGCTGGATGACAAACCAAGC | 3372 |
| Db | 2461 | CTACAAATAAAGGCGAGGATTTCTTGACAGATACACAGAGGCTGCTGGATGACAAACCAAGC | 2520 |
| QY | 3373 | TCTCTTTTCTTTTGTTAACGTGTGCGATGGCGGTGGCTGGCTGTGGCGGTGACCTCATCAG | 3432 |
| Db | 2521 | TCTCTTTTCTTTTGTTAACGTGTGCGATGGCGGTGGCTGGCTGTGGCGGTGACCTCATCAG | 2580 |
| QY | 3433 | CATGGCCCTCATACACACACCGGGGCGTGAATGCTTTATAGCACGGGCGAGATTTCCCC | 3492 |
| Db | 2581 | CATGGCCCTCATACACACACCGGGGCGTGAATGCTTTATAGCACGGGCGAGATTTCCCC | 2640 |
| QY | 3493 | AGCCTATGCGGGTCTCGGCATCTCTTATGCTCCACAGTAAAGGGGGCTTTCAGATTAC | 3552 |
| Db | 2641 | AGCCTATGCGGGTCTCGGCATCTCTTATGCTCCACAGTAAAGGGGGCTTTCAGATTAC | 2700 |
| QY | 3553 | GGTCAGACTGGCATGTGAGACAGAAAGCTGATTCACCTCGGTGAGAGAGATCAATCACTA | 3612 |
| Db | 2701 | GGTCAGACTGGCATGTGAGACAGAAAGCTGATTCACCTCGGTGAGAGAGATCAATCACTA | 2760 |
| QY | 3613 | CATTAAACACTGTGCTTGGAAACGACCTCCGAGAAATTAGAACCAAGGCTCCCTCCCTCGA | 3672 |
| Db | 2761 | CATTAAACACTGTGCTTGGAAACGACCTCCGAGAAATTAGAACCAAGGCTCCCTCCCTCGA | 2820 |

| | | | |
|----|------|--|------|
| QY | 3673 | CTGGCCCCAGAGGAGGAGGTACCTTTTGAGAACCCAGCATAGAGTACCCAGAAAACCT | 3732 |
| Db | 2821 | CTGGCCCCAGAGGAGGAGGTACCTTTTGAGAACCCAGCATAGAGTACCCAGAAAACCT | 2880 |
| QY | 3733 | CCCTCTGCTCTAAAGAAAGTATCCTTCACGATCAACCTAAAGAGAATTTGGCATTTGT | 3732 |
| Db | 2881 | CCCTCTGCTCTAAAGAAAGTATCCTTCACGATCAACCTAAAGAGAATTTGGCATTTGT | 2940 |
| QY | 3793 | GGGGGGAGACAGATACAGGAAGTCCCTGGCTGGGGATGGGCCCTTTCCTGTCGGAGATT | 3852 |
| Db | 2941 | GGGGGGAGACAGATACAGGAAGTCCCTGGCTGGGGATGGGCCCTTTCCTGTCGGAGATT | 3000 |
| QY | 3853 | ATCTGAGAGCTGCATCAAGATTGATGGAGTGAATTCAGTATATTGGCTTTCGCAGCT | 3912 |
| Db | 3001 | ATCTGAGAGCTGCATCAAGATTGATGGAGTGAATTCAGTATATTGGCTTTCGCAGCT | 3060 |
| QY | 3913 | CCGAAGCAAACTCTCTATCATTTCTCCACAGAGCCGGTGGCTGTTCAGTGGCACTGTCAATC | 3972 |
| Db | 3061 | CCGAAGCAAACTCTCTATCATTTCTCCACAGAGCCGGTGGCTGTTCAGTGGCACTGTCAATC | 3120 |
| QY | 3973 | AAATTTGGACCCCTTCAACACAGTACACTGAAGACACAGATTTGGATGGCCCTGGAGAGAC | 4032 |
| Db | 3121 | AAATTTGGACCCCTTCAACACAGTACACTGAAGACACAGATTTGGATGGCCCTGGAGAGAC | 3180 |
| QY | 4033 | ACACATGAAAGAATATTTGCTCAGCTACCTCTGAACTTGAACTGAATCTGAATGAGAGAA | 4092 |
| Db | 3181 | ACACATGAAAGAATATTTGCTCAGCTACCTCTGAACTTGAACTGAATCTGAATGAGAGAA | 3240 |
| QY | 4093 | TGGGATATACCTTCAGTGGGGGAAAGGACAGCTGTTGGATAGCTTAAGCCCTGCTCCG | 4152 |
| Db | 3241 | TGGGATATACCTTCAGTGGGGGAAAGGACAGCTGTTGGATAGCTTAAGCCCTGCTCCG | 3300 |
| QY | 4153 | CCACGTATAGATTCTGATTTTATAGATGAAGCCACAGCTCCCATGGAACAGACAGACTT | 4212 |
| Db | 3301 | CCACGTATAGATTCTGATTTTATAGATGAAGCCACAGCTCCCATGGAACAGACAGACTT | 3360 |
| QY | 4213 | ATTGATTCAAAGAGACCATTCGGAAGACATTTTGACAGCTGTACCATGTGACCATTTGCCCA | 4272 |
| Db | 3361 | ATTGATTCAAAGAGACCATTCGGAAGACATTTTGACAGCTGTACCATGTGACCATTTGCCCA | 3420 |
| QY | 4273 | TGCGCTGCACAGGTTCTAGGCTCCGATAGGATTTATGCTGTGGGCCGAGGACAGGTGGT | 4332 |
| Db | 3421 | TGCGCTGCACAGGTTCTAGGCTCCGATAGGATTTATGCTGTGGGCCGAGGACAGGTGGT | 3480 |
| QY | 4333 | GGAGTTTGACACCCCATCGGTCTCTGTCCAAAGACAGTTCCCGATTCTATGCCATGTT | 4392 |
| Db | 3481 | GGAGTTTGACACCCCATCGGTCTCTGTCCAAAGACAGTTCCCGATTCTATGCCATGTT | 3540 |
| QY | 4393 | TGCTCTGCAGAGCAAGGTGGCTGTCAAGGGCTGCATCCCTGTTTGAAGAAAGTTC | 4452 |
| Db | 3541 | TGCTCTGCAGAGCAAGGTGGCTGTCAAGGGCTGCATCCCTGTTTGAAGAAAGTTC | 3600 |
| QY | 4453 | TTTTCTTTAGACATTTGCATTCCTCGCTGGGGGGGCCCCCTTCATAGCGTCTCTAC | 4512 |
| Db | 3601 | TTTTCTTTAGACATTTGCATTCCTCGCTGGGGGGGCCCC - TCATGCGTCTCTCTAC | 3659 |
| QY | 4513 | CGAAACCTTGCCTTCTCGATTTTATCTTTGGACACAGAGTCCGAGATTTGGCTGTGGT | 4572 |
| Db | 3659 | CGAAACCTTGCCTTCTCGATTTTATCTTTGGACACAGAGTCCGAGATTTGGCTGTGGT | 3719 |
| QY | 4573 | TTTCACCTTTAGGAGAGTCATATTTTGATTTATTTGATTTATTCATTCATGTAAACAA | 4632 |
| Db | 3720 | TTTCACCTTTAGGAGAGTCATATTTTGATTTATTTGATTTATTCATTCATGTAAACAA | 3779 |
| QY | 4633 | AATTTAGTTTTGTTCTTAATTGCACTCTAAAGGTTCAAGGACCGTTATTAATTGT | 4692 |
| Db | 3780 | AATTTAGTTTTGTTCTTAATTGCACTCTAAAGGTTCAAGGACCGTTATTAATTGT | 3839 |
| QY | 4693 | ATCAGAGGCTTAATGAAGCTTTTACGTGAGCTATATCTATTAATTAAATTCGTACAT | 4752 |
| Db | 3840 | ATCAGAGGCTTAATGAAGCTTTTACGTGAGCTATATCTATTAATTAAATTCGTACAT | 3899 |
| QY | 4753 | AGCCATATATTTACATGAAATGTAGCTGTTATTTATTTAAATTAAGACCTGTGCT | 4812 |

D 931 CTTGAAGCTAAAGACATTAAGAAAAAGTCCCTGGGGAGGCTTATCAATATCTGTCCAA 990
Q 961 CGATGGCAGAGAAATGTTTGAAGCAGACGCCGTTGGACGCTGCTGGCTGGAGACCCGT 1020
D 991 CGATGGCAGAGAAATGTTTGAAGCAGACGCCGTTGGAGCAGCTGCTGGCCGAGAACCCGT 1050
Q 1021 TGTGGCATCTTGGGCAATGATTTATATGTAATTTCTGGGACCAACAGGCTTCGCGG 1080
D 1051 TGTGGCATCTTGGGCAATGATTTATATGTAATCTAGAGACCAACGCGCTTCGCGG 1110
Q 1081 ATCAGCTGTTTTATCCCTCTTTTACCCAGAAATGATTTTGCATCAGCGCTCAGACATA 1140
D 1111 ATGGCGCTCTTATCCCTCTTTTATCCGGCAATGATGTTGTCAGAGGCTAACCGGCTA 1170
Q 1141 TTTTCAGAGAAAATGCGTGGCCGACGATGAACGTGTCCAGAAATGAATGAATGTTCT 1200
D 1171 TTTTCAGAGAAAATGCTGTACCGCCAGACGACGCTGTCCAGAAATGAATGAATGTTCT 1230
Q 1201 TACTTCAATTAATTTTCAAAAATGATATGCTGGTGCAAGCAATTTTCTCAGAGTTCA 1260
D 1231 TACCTCAATTAATTTTCAAAAATGATATGCTGGTGCAAGCAATTTTCTCAGAGTTCA 1290
Q 1261 GAAATCCGAGAGAGAGAGCGTGGATTTGAAAAAGCGGCTACTTCCAGACATCAC 1320
D 1291 AAAAATCCGAGAGAGAGAGCGTGGATCTCGAGAAAGCGCGGTACTTTTCAGACATCAC 1350
Q 1321 TGTGGTGTGGCTCCACTGTGTGTGTATGCCAGCGGTGGTGAACCTTCTCTTCATAT 1380
D 1351 CGTGGAGTGTGGCTCCACTGTGTGTGTATGCCAGCGGTGGTGAACCTTCTCTCCACAT 1410
Q 1381 GACCCGTGGCTGTGATGTGACAGACAGACAGCGTTTCACAGTGTGACATCTTCATTC 1440
D 1411 GACCCGTGGCTGTGACAGAGAGAGCGGCTTCACAGTGTGACATCTTCATTC 1470
Q 1441 CAGTACTTTGCTTTGAAATTAACACGTTTTCAGTAAAGCCCTTCAGACAGCTCAGT 1500
D 1471 CAGTACTTTGCTTTGAAATTAACAGTGTGATCTGAGTAAAGCCCTTCAGAGTCCGT 1530
Q 1501 GCGTGTGTGACAGATTTAAGATTTGTTTCTAATGGAAGAGGTTTCATGATAAAGACAA 1560
D 1531 TGCGTGTGACAGATTTAAGATTTGTTTCTAATGGAAGAGGTTTCATGATAAAGACAA 1590
Q 1561 ACCAGCCAGTCTTCACATTCAGATAGAGATGAAGAAATGCCACCTTGGCATGGGCTCTC 1620
D 1591 ACCGCGAGTCTTCACATTCAGATAGAGATGAAGAAATGCCACCTTGGCATGGGCTCTC 1650
Q 1621 CCACTCCAGTATCCAGAACTCGCCCAAGCTGACCCCAAAATGAAAAAGACAGAGAGGC 1680
D 1651 CCACTCCAGTATCCAGAACTCGCCCAAGCTGACCCCAAAATGAAAAAGACAGAGAGGC 1710
Q 1681 TTTCCAGGGCAGAAAGAGAGAGTGAAGCAGCTGACGCGCATGAGCATCAGGCGGTGCT 1740
D 1711 TTTCCAGGGCAGAAAGAGAGAGTGAAGCAGCTGACGCGCATGAGCATCAGGCGGTGCT 1770
Q 1741 GGGAGAGCAAAAGGGCCACCTCTCTGAGACAGTGAAGAGAGCGGCCCTCCGAAGAGA 1800
D 1771 GGGAGAGCAAAAGAGACCTCTCTCTGAGACAGTGAAGAGAGCGGCCCTCCGAAGAGA 1830
Q 1801 AGAAGGCAAGCAGATCCACCTTGGGCAGCTGCGCTTACAGAGAGACACTGACAGATCGA 1860
D 1831 AGAAGGCAAGCAGATCCACCGAGGAGCATGCGTCTGACAGAGAGACACTGACAGATCGA 1890
Q 1861 TCTGAGATCCAAAGAGGTTAACTGGTTGGAATCTGCGGCAAGTGGGAAAGTGAAGAAC 1920
D 1891 CTTAGATATGAAGAGGCAAACTGGTTGGAATCTGCGGCAAGTGGGAAAGTGAAGAAC 1950
Q 1921 CTTCTCATTTTCCAGCCCTTTTGAAGGCGAGATGACGCTCTGAGGCGAGCATTTGCAAC 1980
D 1951 CTTCTCATTTTCCAGCCCTTTTGAAGGCGAGATGACGCTTTTGAAGGCGAGCATTTGCAAC 2010
Q 1981 TGGAACTTGGCTTATGAGCCAGACAGAGCTGAGATCTCAATGCTACTCTGAGAGACAA 2040
D 2011 TGGGACCTTTGCTTACTGTGGCCCAAGAGGCGCTGATTTCTCAAGCCCAAGCTGAGAGACAA 2070

Q 2041 CATCTCTTTTGGGAAGGAATATGATGAAGAAAGATACACTCTGCTGAAACACTGCTG 2100
D 2071 CATCTCTTTTGGGAAGGAATTTGATGAAGAGATTAACCACTGCTGTAACACTTCTG 2130
Q 2101 CCTGAGGCTGACCTGGGCAATCTTTCCAGAGAGGACCTGACGGAATTTGGAGCGCAGG 2160
D 2131 CTTGGGCTGACCTTTGGGCAATCTTTCCAGAGAGGACCTGACCTGAGATTTGGAGCGCAGG 2190
Q 2161 AGCCAACTGAGCGGTGGGAGCGCCAGAGAGATCAGCTTGGCCGCGCTTGTATGTGA 2220
D 2191 AGCCAACTGAGGTGGGAGCGCCAGCGAATCAGCTTGGCCGCGCTTGTATGTGA 2250
Q 2221 CAGAGCATCTACATCTGTGAGACAGCCCTTCAGTGCCTTAGATGCCATGTGGGCAACA 2280
D 2251 CAGAGCATCTATATCTTGGATGATCCCTTCAGTGCCTTAGATGCCATGTGGGCAACA 2310
Q 2281 CATCTCAATGATGCTATCCGGAACATCTCAATCCCAAGCATTTGTTGTATACCA 2340
D 2311 CATCTTCAACAGTCTATCCGGAAGCTCTCAAGTCTTAAGAGGTCCTGTTTGTACACA 2370
Q 2341 CCAATTACAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
D 2371 CCAATTACAGTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2430
Q 2401 GGAAGAGGACCCATGAGACACTGATGATTTAATGTGATGATGATGATGATGATGAT 2460
D 2431 AGAAGAGGATCCCAAGAGAGGCTGATGATTTAATGTGATGATGATGATGATGATGAT 2490
Q 2461 TAACTGTGCTGGGAGAGACACCGAGTGTGATGATGATGATGATGATGATGATGATGAT 2520
D 2491 TAACTGTGCTGGGAGAGACACCGAGTGTGATGATGATGATGATGATGATGATGATGAT 2550
Q 2521 TTTACAGAAAGATCACAGACAAAGGCTCTTAAACAGAGATCAATTAAGAAAGAAAG 2580
D 2551 TTTAC--AAAATCACAGACAAAGGCGCTTAAGCCAGGCTCAATGAAGAAAGAAAGG 2607
Q 2581 AGTAAAGCCAGAGAGAGAGGAGCTTGTGACGCTGGAAGAGAAAGGAGGCTGATGCGC 2640
D 2608 AGTAAAGCCAGAGAGAGAGGAGGAGCTGCTGAGTGTGAGAGAAAGGAGGAGGCTGCGC 2667
Q 2641 CTGTCAGTATATGCTGCTCATCCAGGCTGCTGGGAGGCGCTTGGCATTCCTGTTAT 2700
D 2668 TTTGTCAGTATATGCTGCTCATCCAGGCTGCTGGGAGGCGCTTGGCATTCCTGTTAT 2727
Q 2701 TATGCGCTTTTCAATGCTGATGATGAGGAGCAGCCCTTCAGACCTGATGTTGAGTTA 2760
D 2728 CATGCTCCTTTCATGCTCAATGATGAGGAGCAGCTTCAGACCTGATGTTGAGTTA 2787
Q 2761 CTGATCAAGCAAGAGAGGAGGAGCAGCTGCTGACGCTGAGAGGAGCAGACCTGATGAG 2820
D 2788 CTGATCAAGCAAGAGAGGAGGAGCAGCTGCTGATGAGGAGGAGCAGACCTGATGAG 2847
Q 2821 TGACAGATGAAGAGCAATCTCATATGACATGATGATGATGATGATGATGATGATGAT 2880
D 2848 TGACAGATGAAGAGCAATCTCTCTGACATGATGATGATGATGATGATGATGATGAT 2907
Q 2881 GGCAGTATGCTGATCTGTAAGAGCATTCGAGAGTGTCTTTGTCAAGGAGCAGCTGCG 2940
D 2908 GGCAGTATGCTGATCTGTAAGAGCATTCGAGAGTGTCTTTGTCAAGGATACATGAG 2967
Q 2941 AGCTTCTCCCGCTGATGATGAGAGCTTTTCCGAAGGATCTTCGAAGCCCTATGAGAT 3000
D 2968 AGCTTCTCCCGCTGATGATGAGAGCTTTTCCGAAGGATCTTCGAAGCCCTATGAGAT 3027
Q 3001 TTTTTCAGAGACCCCAAGGAGATTTCAACAGGTTTTCACAAAGATGATGATGAT 3060
D 3028 TTTTTCAGAGACCCCAAGGAGATTTCAACAGGTTTTCACAAAGATGATGATGAT 3087
Q 3061 TGAGTGTGGGCTGCTTCAAGGCGAGATGTTATCAGAAAGCTTATCTGCTGCTTCT 3120
D 3088 GAGCTGCTGCTGCTTCAAGGCTGAGATGTTATTCAGAAATGATGATGCTGCTGCTTCT 3147

| | | | |
|----|------|--|------|
| QY | 3121 | CTGTGGGAATGATGGAGGAATCTCCCGGGTTCCTGTGGCAGTGGGGCCCTTGT | 3180 |
| Db | 3148 | CTGTGGATATATGGACGGGGCTCTTCCATAGTTCTTGTGGAGGGGGCCCTTCT | 3207 |
| QY | 3181 | CATCCTCTTTTCAAGTCTCGACATTTCTTCAGAGGTCCTGATTGGAGCTGAAGCCT | 3240 |
| Db | 3208 | CATCCTCTTCTCAGTTCTACACTTGTCTTCAGAGGTCCTGATTGGAGGCTGAAGCCT | 3267 |
| QY | 3241 | GGACATATATCAGCAGTCACTTTCCTCCACATTCAGCTCCAGCATACAGGGCCTTGC | 3300 |
| Db | 3268 | GGACATATATCAGCAATCTCTTTCCTCCCATATCAGTCCAGCATTCAGGGCCTGCG | 3327 |
| QY | 3301 | CACCATCAGCCTACAAATTAAGGGAGAGATTCTGCAGATACAGAGAGCTGTGGA | 3360 |
| Db | 3328 | CACCATCAAGCCTAATAACAAAAGGAGAGATTCTACACAGTATCAGAGAGCTTGTGA | 3387 |
| QY | 3361 | TGACAAACAAGCTCTCTTTTCTTTTGTATTAGTGGGATGGGGTGGCGGTGTGGCGCT | 3420 |
| Db | 3388 | TGACAAACAAGCCCCCTTCTTCTTGTTCACCTGTGCATAGAGTGGCTGTGGAGTGGCCT | 3447 |
| QY | 3421 | GGACCTCATCAGCATCGCCCTCATCCACCAAGGGGGTGAATGATCGTTCTTATGACAGG | 3480 |
| Db | 3448 | GGACCTCATCAGCATCGCCCTGATACCAACAAGGGGCTCATGATTGTTCATATGACAGG | 3507 |
| QY | 3481 | GCAGATTCGCCCAAGCCTATGCGGGGTCTGCCATCTCTTATGTCTGCCAGTTACGGGCT | 3540 |
| Db | 3508 | GCAGATTCCTTTCAGCTAGCAGCGGGGCTCCGCCATCTCTACGCTGTGCAGTTAACCGGACT | 3567 |
| QY | 3541 | GTTCACATTTACGGTTCAGATCTGCATCTGAGACAGAAAGTGCATTCACTCGGTGGAGAG | 3600 |
| Db | 3568 | CTTCACATTTTACGGTTAGACTGGCGTGGAGACAGAGACAGCGTTCACTCAAGTGGAGAG | 3627 |
| QY | 3601 | GATCAATCACTTACATTAAAGCTGTGCTTTGGAAGACACTGGCAGAAATTAAGACAAAGC | 3660 |
| Db | 3628 | GATCAACACTTACATAAAGACTCTCTTTGGAGAGACTGGCAGAAATTAAGACAAAGC | 3687 |
| QY | 3661 | TCCCTCCCTGACTGCGCCCAAGAGGAGAGGTGACTTTTGAAGACGAGAGATGAGTA | 3720 |
| Db | 3688 | TCCCTCCCATGACTGGGCCCAAGAGGAGAAATAACTTTGAGATGCGGAATAGATA | 3747 |
| QY | 3721 | CCGGAAGAAACCTCCCTCTCTCTTAAGAAAGTATCTTACAGATCAACAACCTAAAGAA | 3780 |
| Db | 3748 | CCGGAAGAAATCTCCCTCTGCTCTTAAGAAAGTCTTACACATCAAGCCCAAGGAAAA | 3807 |
| QY | 3781 | GATTGGCATTTGTGGGGGCGGACAGATACAGAGAAAGTCTTGGCGGGGATAGGCCCTTTCG | 3840 |
| Db | 3808 | GATAGGCAATTTGTGGAGGAGAACAGAGTCAAGGAAAGTCTTCAAGGAGATAGGCCCTTTCG | 3867 |
| QY | 3841 | TCGTGTGGAATTATCTGAGGCTGCATCAAGATTGATGAGTGAAGATCACTGATATTGG | 3900 |
| Db | 3868 | CCTGTGTGGAATTATCTGAGGCTGCATCAAGATCGATGAGTGAAGATCACTGATATTGG | 3927 |
| QY | 3901 | CCTTGGCCGACTCCGAAGCAAACTCTCTATCATCTCTCAAGAGCGGGGTGTTCAGTGG | 3960 |
| Db | 3928 | CCTTGGCCGACTCCGAAGCAAACTGCATCATCTCTCAAGAGCCAGTGTCTTCAAGGG | 3987 |
| QY | 3961 | CACGTGCAGATCAAAATTTGGAACCCCTTCAACAGATACATGAAAGCAGATTTGGAGTGC | 4020 |
| Db | 3988 | CACAGTGCAGATCAAAACTGAGCCCTTCAACAGATACAGAGAGCAGATCTGGAGTGC | 4047 |
| QY | 4021 | CCTGGAGAGACACACATGAAGAATGATGTCTCACTACACTCTGAACCTTGAATCTGA | 4080 |
| Db | 4048 | CCTGAGAGAAACACACATGAAGAATGTATGTGCCAGTACTCTGAACCTGGAATCTGA | 4107 |
| QY | 4081 | AGTATGAGAAATGSGGATTAATCTTCTCAGTGGGGAGACGGCCTTGTGATAGCTAG | 4140 |
| Db | 4108 | AGTATGAGAAAGCGGACAACTTCTCTGTGGGGAGACGGCAGTTGTATGATAGCAAG | 4167 |
| QY | 4141 | AGCCCTCTCGGCACAGTGAAGATTGTGATTTAGATGAAGCAGACAGTGCATGAGAC | 4200 |
| Db | 4168 | GGCCCTCTCAGTCACTGTAAGATTCTGATTTTAAATGAAGCTACAGCCGCTATGAGAAC | 4227 |
| QY | 4201 | AGAGACAGACTTATTGATTCAGAGACCATCGAGAGCACTTTGGACACTGTACCATCT | 4260 |

| Db | 4228 | AGAGCAGACTTACTGATTCACAGAGACCATTCGGAGACATTTCGTGACTGCACCATGCT | 4287 |
|----------------------------|---|---|----------------------|
| QY | 4261 | GACCAATGCCCATCCCTGCACACACGGTTCCTAAGCTCCGATAGATTATGTCGGCCCA | 4320 |
| Db | 4288 | GACCAATGCCCATCCCTGCACACACAGTTCCTGAGCTCTGATAGATCATGTCGTGGCCCA | 4347 |
| QY | 4321 | GGGACAGGTGCTGAGATTGACACCCCATCCGCTCTTCTGTCCAAACAGACAGTTCGCCATT | 4380 |
| Db | 4348 | GGGACAGGTGCTGAGATTGACACCCCATCCGCTCTTCTGTCCAAACAGACAGTTCGCCATT | 4407 |
| QY | 4381 | CTATCCCAATGTCGTGCTGCAGAGAACAAAGGTGCTGTCAAGGGCTGACCTCCCTCGT | 4440 |
| Db | 4408 | CTATCCCAATGTCGTGCTGCAGAGAACAAAGGTGCTGTCAAGGGCTGACCTCTCC--- | 4464 |
| QY | 4441 | TCACCAATGTCGTGCTGCTTTAGAGCATTCGCCATTCCTGCCTGGGGGCGCCCTTCATC | 4500 |
| Db | 4465 | ACGCTGGAAGTCTCTCTCTCTGAGAGGCTTGCCATTCCTCTCGTGGGCGCTCATCTC | 4523 |
| QY | 4501 | GGCTCCCTCCACGAAACCTTCGCTTCCTCGATTTATCTTCCGACACAGCTCCGAT | 4560 |
| Db | 4524 | -----GTCCTGCTGAACCTCCCTCTCTCTGCTT-TCCTCTCACGACAGTTCAGGGT | 4576 |
| QY | 4561 | TGGCTGTGTCGTTCACCTTTAGGAGAGTCATATTTGATATGTAATTTATTCGATAT | 4620 |
| Db | 4577 | TAGC-TGTGTGTCACCTTTAAGGAAATCATATTTGATATTTGTAATTTATTCCTAT | 4635 |
| QY | 4621 | TCATGTAACAAATTTTGGTTTGTCTTATTCACATC-TAAAGGCTCAGGAAACG | 4679 |
| Db | 4636 | TCATGTAACAAATTTTGGTTTGTCTTATTCACATC-TAAAGGCTCAGGAAACG | 4695 |
| QY | 4680 | TTATTTATTAATGTTATCAGAGGCTTAATGAAGCTTTATACGTAGCTATATCATATA | 4739 |
| Db | 4696 | TTACATATA-CGTACAGAGGCTTAATGAAGCTTTATACGTAGCTATATCATATA | 4754 |
| QY | 4740 | TAAATCTGACCTACCTATATTTACAGTGAAGAGTAAGCTGTTATTTATATATAA | 4799 |
| Db | 4755 | TAAATCTGACCTACCTATATTTACAGT-AAAATGAAGTGGTTTATTTCTATTTAAAG | 4812 |
| QY | 4800 | TAAAGCTGCTGCTAAAAA 4818 | |
| Db | 4813 | CGAGCAGCTGCTAAACACA 4831 | |
| RESULT | 10 | | |
| LOCUS | A64695 | 1761 bp | DNA |
| DEFINITION | Sequence 61 from Patent W09731111. | | PAT |
| ACCESSION | A64695 | | 29-MAR-1999 |
| VERSION | A64695.1 | GI:4530760 | |
| KEYWORDS | | | |
| SOURCE | unidentified. | | |
| ORGANISM | unclassified. | | |
| REFERENCE | 1 (bases 1 to 1761) | | |
| AUTHORS | Oude,E.R., Paulsma,C.C., Bosma,P.J., Borst,P., Evers,R., Kool and Marcel. | | |
| TITLE | A FAMILY OF ORGANIC ANION TRANSPORTERS, NUCLEIC ACIDS ENCODING THEM, CELLS COMPRISING THEM AND METHODS FOR USING THEM | | |
| JOURNAL | INTROGENE BV (NL) | | |
| COMMENT | other publication AU 1736697 19970910. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..1761 | | |
| BASE COUNT | 427 a 448 c 469 g 413 t | | 4 others |
| ORIGIN | | | |
| Query Match | 34.5%; | Score 1674.4; | DB 9; Length 1761; |
| Best Local Similarity | 99.5%; | Fred. No. 0; | |
| Matches 1722; Conservative | 0; | Mismatches | 1; Indels 7; Gaps 4; |

| | Matches 2108; | Conservative | 0; | Mismatches 1899; | Indels 125; | Gaps |
|----|---------------|--|----|------------------|-------------|------|
| QY | 284 | GAGGCGCTCTCTTGATGCTCCATGCATTCCTCAGCTCAGCATCTCGATGATGAGGACAT | | | | 343 |
| Db | 217 | GATGGCCCCCTGGAGTCACCAAGAGAAATCTTGAGGCTCCAGGGAGGGCACCTGTCCCA | | | | 276 |
| QY | 344 | CCCAAGGGAAGTACCATCATGCGCTTGAATGCTGGAAGCCCATCG-----GACTACT | | | | 397 |
| Db | 277 | CCGTGGGGGAAGTATGATGCTGCTTGGAGAACCATGATTTCCCTTCCGTCCTCAAGCCGAGG | | | | 336 |
| QY | 398 | TGCAAAACACAGCACCCAGTGGACATGCTGGGCTTTTTCCTGTATGACTTTTTCGTGG | | | | 457 |
| Db | 337 | TTTCCCTGCCCCCACCACCCCTGGACATGCTGGCTCTGTCTCCTACCTCAACCTGTCATGG | | | | 396 |
| QY | 458 | CTTTCCTTCGTGGCCCGGTGGGCCCAACAAGAGGGGGAGCTTCATATGGAAGA -GCTGG | | | | 516 |
| Db | 397 | CTCAC-----CCGCTCATGATCCAAAGCTTACGAGTGGCTGTAGATGAGAACCACTATCCC | | | | 452 |
| QY | 517 | GTCCTGTCCAAAGCAGAGTCTTCTGAGGTGAACTCGAGAAACACTAGAGAGACTGTGGCA | | | | 576 |
| Db | 453 | TCCACTGTGACATGCATGTATGCCCTCAGACAAAATGTCCAAAGGCTTACCGGCTTTGGGA | | | | 512 |
| QY | 577 | AGAAAGCTGAATGAATGGGGCCACAGACGCTCTTCCCTGCCAAGGCTTGTGTGATCTT | | | | 636 |
| Db | 513 | AGAAAGTCTTCMAAGCGAGGAGTGAATAAAACCTTCAGGCTTCGTGTATCTGTAGGTT | | | | 572 |
| QY | 637 | CTGGCGCACACAGGCTCATCTCTCCATCGTGTGCGCTGATGATCAGACAGCTGGTGGCTT | | | | 696 |
| Db | 573 | CCAGAGAACAAAGTTGATTTTCGATGACACTCTTGAGGACATCTGCTTCGTATGGCCAGTGT | | | | 632 |
| QY | 697 | CAGTGGACACGCTTCATATGTTGAACACCTCTGTGAGTATACCCAGGACAGAGATCTTA | | | | 756 |
| Db | 633 | ACTCGGGCCAAATATGATATTAACCAAGATCTCTGAAATTTTCAGAAAGACAGTGGGGAA | | | | 692 |
| QY | 757 | CTGTGAGTACAGCTTGTATTAGTACTGGGCTCCCTCCGACGAGAAATGTGCGGCTTGA | | | | 816 |
| Db | 693 | TGTTTCCATGAGTAGTGGGACCTGCTGGCTTTTCTCTCGCATGTGTGAAGCTCTCT | | | | 752 |
| QY | 817 | GTCGCTTGCACTGACTTGGGGCAATTGAATTAACGAACCGGTGTCCGCTTGGGGGGCCAT | | | | 876 |
| Db | 753 | GAGTTTCTCCTCCAGTGTGATCATCAACCACAGCACAGCATCAGGTTCCGAGCACGCTGT | | | | 812 |
| QY | 877 | CCTAACCATGGCAATTAAAGATGCTTAAGTTAAATTAAGAAACATTAAAGAAATCCCTGGG | | | | 936 |
| Db | 813 | TTTCCCTCTTTCCTTTGAGAACTCATCCAAATTTAAGTGTGTAATATACATCACTCAGG | | | | 872 |
| QY | 937 | TGAGCTCATCAACATTTGCTCCACAGATGGGCGAGAAATGTTTGAAGCAGACCGCTTGG | | | | 996 |
| Db | 873 | AGAGGCACTACACTTCTTCCACCGGATGTAAACTACGTTTGAAGGGGTGTGCTATG | | | | 932 |
| QY | 997 | CAGCGCTGCGCTGGAGGACCCGTTGTTCCTCATCTTAGCATGATTTATTAATGTATTT | | | | 1056 |
| Db | 933 | ACCCCTAGTACTGATCAACCTGGGCATCGCTGGTCACTGACACATTTCTCTACTTCAT | | | | 992 |
| QY | 1057 | TCTGGACCAACAGGCTCTCCCTGGGATCAGACTTTTATCTCTTTTATCCAGCAATGAT | | | | 1116 |
| Db | 993 | TATTTGGATACATGATGATTTATTTGCCATCTTATGCTATCCCTGGTTTTTCCACTGGAGGT | | | | 1052 |
| QY | 1117 | GTTTGCATCAGCGCTCACAGCATTTTTCAGGAAAAATCGTGGCCGACAGGATGACG | | | | 1176 |
| Db | 1053 | ATTCTATGACAGAAATAGGCTGTGAAGGCTCCAGCATCACACATCTGAGAGTCAAGGACCGAGG | | | | 1112 |
| QY | 1177 | TGTCCAGAAAGATGAAATGAACTTCACTTACCATTAATTAATTAACAAATGATAGCCGGGT | | | | 1236 |
| Db | 1113 | CATCCGTGTGACAGATGAAATTTCTCCTTGCACTTAAGCTGATTAATAATGTACATAGGGA | | | | 1172 |
| QY | 1237 | CAAAACATTTTCTCAGAGTGTTCAGAAAAATCCGCGAGAGAGGAGCTCGGATATTGGAAAA | | | | 1296 |
| Db | 1173 | GAAACCATTTTGCAAAATATCATTTGAAGACTTAAGAAGAGGAAGAAAGAACCTATTGGAGAA | | | | 1232 |
| QY | 1297 | AGCGGGATCTTCCAGAGCATCATCTGTGGGTGTGGCTCCCATTTGTGTGTGATTTCCAG | | | | 1356 |
| Db | 1233 | GTGCGGGCTGTTCGACAGAGCTCAGAGTATTAACCTTTGTTCATCATCCCCGACAGTGGCAC | | | | 1292 |

| | | | |
|----|------|--|------|
| Oy | 1357 | CGTGTACACTTCTCTCTGTATATATGACCCCTGGGCTTCATCTGCAGACGACACAGCTTT | 1416 |
| Oy | 1357 | | |
| Db | 1293 | AGCGGCTGGGGTTCTCAATCCACACATCTTAAAGCTGAAGCTCACAGCTCAATGGCTTT | 1352 |
| Oy | 1417 | CACAGTGTGACATCTCTTCAATTCATAGCATTTTGTGTAAGTAAACCCGTTTCAGT | 1476 |
| Oy | 1417 | | |
| Db | 1353 | CAGACTGCTAACCTCCCTTGAAATCCCTCGCGCTGTGAGTGTTCTTTGTGCCATTTGACGT | 1412 |
| Oy | 1477 | AAAGTCCCTCTCAGACAGCCCTACGTGGCTGTTGACAGATTTAAAGATTGTTCTTATATGA | 1538 |
| Oy | 1477 | | |
| Db | 1413 | CAAAAGTCTCCGAAATTCCAAAGTCTGAGTGAAGAGATTCAAGAAAGTTTCTCCAGGA | 1472 |
| Oy | 1537 | AGAGGTTACACTGATTAAGAACAACAGCAGCATCTCCATCAATCAAGATAGAGATGAATAA | 1586 |
| Oy | 1537 | | |
| Db | 1473 | GAGCCCTGTTTTTATATGTCAGACATTTACAAAGACCCACCAAGCTCTGTGTTGABGA | 1532 |
| Oy | 1597 | TGCACACTTGGCATGGGACTCTCCCACTCCATATCCAAACTGCCCAAGCTGACCC | 1656 |
| Oy | 1597 | | |
| Db | 1533 | GGCCACCTTGTCATGGCACAGACCTGTCCGGATC----- | 1566 |
| Oy | 1657 | CAAAATGAAAAAGACAAAGAGGCTTCCAGGGGCAAGAAAGAAAGTGAAGCAGCTGTA | 1716 |
| Oy | 1657 | | |
| Db | 1570 | ----- | 1568 |
| Oy | 1717 | GCGCACGTAGCATCAGCGGTTGCTGGCAGAGCAAGAAAGCCACTCTCTGTGACAGTGA | 1776 |
| Oy | 1717 | | |
| Db | 1570 | -----GTCAATGGGCACTGTGAGCTGGAGAGAAACGGGACATGCTTGTGAAGGGATGAC | 1622 |
| Oy | 1777 | CGAGCGCCCACTGCCAAGAGAAAGCAAGCAACATCCACTCTGGGCACTCTGCGCTT | 1836 |
| Oy | 1777 | | |
| Db | 1623 | CAGGCTTAGAGATCCCTCGGGGCAAGGAAAGGAAACAGCCTGGGCGCCAGAG---- | 1677 |
| Oy | 1837 | ACAGAGGACACTGCACAGCATCGATCTGAGATCCAAAGAGGGTAAACTGTTGGATCTG | 1896 |
| Oy | 1837 | | |
| Db | 1678 | -----TTGCACAAAGATCAACCTGGTGGTGTCCAAAGGGGATGATGTAGGGGTCTG | 1727 |
| Oy | 1897 | CGCGAGTGTGGGAAGTGGAAAAACCTCTCATATTTAGCCATTTTAGCCCAATATACCT | 1956 |
| Oy | 1897 | | |
| Db | 1728 | CGCGAACACGGGGAGTGTAAAGCAACCTGTTGTCAAGCATCTCGAGAGATCATCATT | 1787 |
| Oy | 1957 | TCTTAGAGGCGAGCATTCATCAATGATGAACCTTCGCTTAATGTGGCCACAGCCCTGAT | 2018 |
| Oy | 1957 | | |
| Db | 1788 | GCTGAGAGGCTCGGTTGGGGGTGCAGSBAACCTGCGCTATGTCCCAACAGCCTGGAT | 1847 |
| Oy | 2017 | CCTCAATGCTACTGTGAGAGCAACATCTGTTTGGGAAGGAATATGTAAGAAAGATA | 2076 |
| Oy | 2017 | | |
| Db | 1848 | CGTCAGCGGGAACATCAAGGGAAGAACTCTCATGGGAGGCGCTATATGACAAAGGCCGATA | 1907 |
| Oy | 2077 | CACCTCTGTGTAACAGCTGTCTCCGABGCTGACCTGGCCATCTTCCAGCAGCAGGA | 2136 |
| Oy | 2077 | | |
| Db | 1908 | CCTCAGGTGCTCACTGCTGTCTCTCGAATCGGAGCTGTGAACCTTCTCCCTTGGAGA | 1967 |
| Oy | 2137 | CCTGACGAGATTTGGAAGAGGAGAGCACTGCAGCGGTGGGACGCGCACAGAGATAG | 2196 |
| Oy | 2137 | | |
| Db | 1968 | CATGACAGAAATTTGGAAGAGGAGGCGCCCAACCTCTCGGGGGGCGAAGAACAGAGATCAG | 2027 |
| Oy | 2197 | CCTTGGCCGGGCTTTGTAATGATGACAGAGACATCTATCATCTGTGAGAGACCCCTCAATGC | 2256 |
| Oy | 2197 | | |
| Db | 2028 | CCTGGCCCGGCGCTCTTAATTCGACCTCAGATCTACTCTGTGAGAGACCCCTCTGTGC | 2087 |
| Oy | 2257 | CTTAGATCCCATGTGGGCAACCAATCTTCAATAGTGCATCCGGAACATCTCAAGTC | 2316 |
| Oy | 2257 | | |
| Db | 2088 | TGTGAGCGCCCACTGGGGGAAGCAATTTTGGAGAGTGCATTTAAGAAACACACTCAGGGG | 2147 |
| Oy | 2317 | CAAGACACTTCTGTTTGTATACCAACAGTACAGTACCTGTGACTGTGATGAAGTGT | 2376 |
| Oy | 2317 | | |
| Db | 2148 | GAAGACGCTGTCCAGGAGGACCCACAGCTGCAATCTTGAATTTTGTGGCCAGATCAT | 2207 |
| Oy | 2377 | CTTCAATGAAGAGGCTGTATATACGGAAGAAAGCACCCATGAGAGACCTATGAATTTTAA | 2436 |
| Oy | 2377 | | |
| Db | 2208 | TTTGTGTGAAAAATGGAAAAATCTGTGAAATGGAACCTCACAGTGAATTAATACAGAAAA | 2267 |
| Oy | 2208 | | |

| | | | |
|---|------|---|------|
| D | 2142 | CACATTACCTGGGGCCA--GGACGACGCTCCACACATGAAAGCATCACTTCTCATC | 2199 |
| Q | 1871 | CMAAGGGTAACCTGGTTGGAAATCTGGCGCAGTGTGGGAATGGAAAACTCTCTCAT | 1930 |
| D | 2200 | CCCGAAGGTGCTTTTGGGCGCGTGGGGCCAGAGTGGGGCTGGGAAATCTCTCCCTGCTC | 2259 |
| Q | 1931 | TCAGCCATTTTAGCCAGATGACGCTTCTAGAGGGCAGATTGCATGAGTGGAACTTC | 1990 |
| D | 2260 | TCACCCCTCTGGCTGTGAGTGGACAAAGTGGAGGGGCACTGGCTATCAAGGGCTCCGTG | 2319 |
| Q | 1991 | GCTATGTGGGCCAGCGGCTGGATCTCAATGCATCTCAGAGCAACATCTCTGTT | 2050 |
| D | 2280 | GGATGTACGCTGGAGGAACCATATTACAGTCCGTGATACAGGCTGTGGCCCTCCCA | 2439 |
| Q | 2111 | GACCTGGCCATTCTTCCACAGCCACTGACGGAGATTGGAGACGAGAACCTG | 2170 |
| D | 2440 | GACCTGGAATCTGTGCCCACTGGGATCGGACAGATGTGGCAAGAGGGGTGAACTG | 2499 |
| Q | 2171 | AGGGGTGGGAGCGCCAGAGAGATCAACCTTGGCCGGGCTTTTATAGTACAGAGCATC | 2230 |
| D | 2500 | TCCTGGGGACAGAAAGACGCGCTGAGCCTGGGCCGGGCGCTACTCCAGCGTGCAT | 2559 |
| Q | 2231 | TACATCCTGGACGACCCCTCACTGCTTAGATGCCCATGTGGGCAACACATCTTCAT | 2290 |
| D | 2560 | TACCTTTCGATGATCCCTCTCAGCAGTGAATGCCCATGTGGAAAAACATCTTTGAA | 2619 |
| Q | 2291 | AGTCTAT-----CGGAAACATCTCAAGTCAGACAGATTTCTGTTTATCCACCAG | 2344 |
| D | 2620 | AATGTGATTTGGCCCCAAGGGGATGCTGAAGAACACAGACCGGATCTTGGTACGCGACG | 2679 |
| Q | 2345 | TTTACAGTACCGGTGTGACTGTGATGAAGTATCTTATGAAGAGGGCGTATTACGGA | 2404 |
| D | 2680 | ATGAGCTACTTGGCCGAGGTGGAGCGTATCATGTATGTAGTGGCGGCAAGATCTGTAG | 2739 |
| Q | 2405 | AGAGGCAACCCATGGAGAAATGATGAATTTAAATGGTACATGCTACCATTTTAAATAC | 2464 |
| D | 2740 | ATGGGCTCTTACACAGAGAGTGTGCGGTCCGAGAGGGGCGCTTGCGTGAATCTGTGATCC | 2799 |
| Q | 2465 | CTGTGCTGGGAGACACCGCAGTT---GAGATCAATTCAAAAAGAAACACAGTGT | 2521 |
| D | 2800 | TATCCGACAGACGAGCGAGGAGCATGTGACAGAGAAAGGGGGTCAAGGGGCTCAGCGT | 2859 |
| Q | 2522 | TCACAGAAAGTCAACAAGCAGG-----GTCTTAAACAGAGTATTAAGAAAG | 2573 |
| D | 2860 | CCAGGAGGAAGAACGAAAGCAATAGGAATGSCATGTCTGTACGACAGTGCAGGGAAAG | 2919 |
| Q | 2574 | AAAAAGAGTAAAGCCAGAGAGAGGGCAGCTTGTGAGCTG----- | 2615 |
| D | 2920 | CAATCTGAGAGACAGCTCAGCACTCTCTCTCTATAGTGGGGAATACAGACGACACAC | 2979 |
| Q | 2616 | AAGGAAAGGGCAGGGTCTAGTCCCTGTCTCAGTATATGTGTCTACATCAAGCTGCTG | 2675 |
| D | 2980 | AACGACACCGAAGAACTGCAAGAAAGTGAAGGCCAAGAAAGAGAAACCTCGGAAGCTGATG | 3039 |
| Q | 2676 | GGGGC-----CCCTTGGGATCTCTGGTTATATAGGCCCTTTTATGCTGAATGTAGG | 2728 |
| D | 3040 | GAGGCTGACAAGGGCAGACAGAGGGCAGTCAAGCTTTCGTGTACTGGGACTACATGTAG | 3099 |
| Q | 2729 | AGCAGCCGCTTCAGACACCTGTTGGTTGAGTTACTGGATCAACAGAGAAAGCGGAAACAC | 2788 |
| D | 3100 | GCCATTCGGACTTTCATCTCTTCTCAGCATCTTCTTTTATGTGTATACCATGTGTCC | 3159 |
| Q | 2789 | ACTGTGACT-----CGAGGGAACGAGACCTCGGTGAGTGAACAGCATGAAGACATCT | 2842 |
| D | 3160 | GCGCTGCTTCCAACATTTTGGTCTCACGCTCTGGACTGATGACCCCATGTCAAGGGGACT | 3219 |
| Q | 2843 | CATATGAGTACTATGTGCAGACATCTACGCCCTTCTCATBTGGCAGTATCTGATCTGAAA | 2902 |
| D | 3220 | CAGAGACACAGAAAGTCCGGCTGACGCTCTATGAGCCCTTGGGATCTTCAACAGGATC | 3279 |

Tue Nov 13 11:54:57 2001

us-09-528-031-1.rge

Page 33

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2001, 13:04:42 ; Search time 278.09 seconds
(without alignments)
10944.080 Million cell updates/sec

Title: US-09-528-031-1
Perfect score: 4847
Sequence: 1 GGCTACTGCTCGGAGCGTG.....AAAAAAAAAGGCGGCCGC 4847

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq.0601.*
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 4847 | 100.0 | 4847 | 21 | AAAA0481 |
| 2 | 4845.4 | 100.0 | 4847 | 19 | AAV5682 |
| 3 | 4792.6 | 98.9 | 5838 | 20 | AAZ30079 |
| 4 | 4792.6 | 98.9 | 5838 | 21 | AAZ94745 |
| 5 | 4775.4 | 98.5 | 4781 | 22 | AAC85287 |
| 6 | 1674.4 | 34.5 | 1761 | 18 | AAZ94054 |
| 7 | 1054.2 | 21.7 | 2167 | 18 | AAZ94055 |
| 8 | 511.2 | 10.5 | 5011 | 19 | AAV31498 |
| 9 | 511.2 | 10.5 | 5011 | 20 | AAI19818 |
| 10 | 511.2 | 10.5 | 5011 | 20 | AAZ21977 |
| 11 | 511.2 | 10.5 | 5011 | 21 | AAZ94741 |

| | | | | | |
|----|-------|------|------|----|----------|
| 12 | 511.2 | 10.5 | 5011 | 21 | AAZ90193 |
| 13 | 511.2 | 10.5 | 5011 | 21 | AAZ39556 |
| 14 | 506.4 | 10.4 | 4864 | 17 | AAI14911 |
| 15 | 506.4 | 10.4 | 4885 | 17 | AAI14910 |
| 16 | 506.4 | 10.4 | 5011 | 15 | AAQ65377 |
| 17 | 506.4 | 10.4 | 5011 | 17 | AAI17173 |
| 18 | 506.4 | 10.4 | 5011 | 19 | AAV31497 |
| 19 | 506.4 | 10.4 | 5011 | 20 | AAI19817 |
| 20 | 506.4 | 10.4 | 5011 | 20 | AAZ21976 |
| 21 | 506.4 | 10.4 | 5011 | 21 | AAZ60522 |
| 22 | 506.4 | 10.4 | 5011 | 21 | AAZ90192 |
| 23 | 506.4 | 10.4 | 5011 | 21 | AAZ39555 |
| 24 | 437.4 | 9.0 | 4231 | 20 | AAZ30078 |
| 25 | 420.4 | 8.7 | 5889 | 19 | AAV31499 |
| 26 | 420.4 | 8.7 | 5889 | 20 | AAI19819 |
| 27 | 420.4 | 8.7 | 5889 | 20 | AAZ21978 |
| 28 | 420.4 | 8.7 | 5889 | 21 | AAZ90194 |
| 29 | 420.4 | 8.7 | 5889 | 21 | AAZ39557 |
| 30 | 404.4 | 8.3 | 437 | 21 | AAZ02340 |
| 31 | 402 | 8.3 | 4834 | 17 | AAI14912 |
| 32 | 390.4 | 8.1 | 4918 | 18 | AAI14912 |
| 33 | 389 | 8.0 | 4669 | 17 | AAI14913 |
| 34 | 382.2 | 7.9 | 5586 | 18 | AAI14913 |
| 35 | 380.6 | 7.9 | 4864 | 21 | AAZ94737 |
| 36 | 380.6 | 7.9 | 4868 | 19 | AAV26068 |
| 37 | 365.2 | 7.5 | 4931 | 19 | AAZ29584 |
| 38 | 362.4 | 7.5 | 4509 | 20 | AAZ30081 |
| 39 | 356.8 | 7.4 | 4762 | 18 | AAZ94052 |
| 40 | 356 | 7.3 | 5079 | 20 | AAZ30080 |
| 41 | 346 | 7.1 | 4877 | 16 | AAI01569 |
| 42 | 341 | 7.0 | 5110 | 16 | AAI01568 |
| 43 | 340.8 | 7.0 | 463 | 19 | AAV5683 |
| 44 | 340.8 | 7.0 | 463 | 21 | AAAA0482 |
| 45 | 340.8 | 7.0 | 463 | 22 | AAZ85288 |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| ID | AAAA0481 standard; cDNA; 4847 BP. |
| XX | AAAA0481; |
| AC | 23-NOV-2000 (first entry) |
| DT | Human MRP-beta cDNA. |
| XX | |
| DE | Human MRP-beta cDNA. |
| XX | |
| KW | MRP-beta; multidrug-resistance associated protein; human; cytosolic; |
| KW | chemotherapy; cancer cell; drug discovery; cytotoxic; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FT | Key |
| FT | CDS |
| FT | Location/Qualifiers |
| FT | 116..4429 |
| FT | /*tag= a |
| FT | /Product= "MRP-beta" |
| PN | US6077936-A. |
| XX | |
| PD | 20-JUN-2000. |
| XX | |
| PF | 16-APR-1998; 98US-0061400. |
| XX | |
| PR | 16-APR-1997; 97US-0843459. |
| XX | |
| PA | (MILL-) MILLENNIUM PHARM INC. |
| XX | |
| PI | Shyjan A; |
| XX | |
| DR | WPI; 2000-430613/37. |
| DR | P-PSDB; AAB10225. |

Multidrug resistanc
Human MRP variant
cDNA encoding mult
cDNA encoding mult
Multidrug resistanc
cDNA encoding mult
Human multidrug re
Human multidrug re
Human multidrug re
Multidrug resistanc
Human multidrug re
Human multidrug re
cDNA encoding a hu
Murine multidrug r
Mouse multidrug re
Mouse multidrug re
Murine multidrug r
Human secreted pro
cDNA encoding mult
Rat canalicular mu
cDNA encoding mult
Human canalicular
Human ATP binding
Human secreted pro
Human canalicular
Homo sapiens sulph
cDNA encoding a hu
Human multidrug re
cDNA encoding a hu
Hamster sulphonylu
Rat sulphonylurea
Human multidrug re
Human MRP-beta cDN
MRP-beta probe, fo

Model multidrug-resistance associated polypeptide useful for improving the effectiveness of a chemotherapeutic regimen to eradicate multidrug-resistant transformed cells especially cancer cells -

Claim 6; Fig 1A-D; 43pp; English.

This invention describes a novel model multidrug-resistance associated polypeptide, MRP-beta (1) which has cytostatic activity. Inhibitors of (1) are useful for improving the effectiveness of a chemotherapeutic regimen to eradicate multidrug-resistant transformed cells, especially cancer cells, from the body of a mammal, preferably human. (1) is also useful for drug discovery, especially to the design of novel chemotherapeutic drugs that are cytotoxic to cells expressing (1). This sequence encodes the human MRP-beta protein which is described in the method of the invention.

Sequence 4847 BP; 1226 A; 1178 C; 1263 G; 1180 T; 0 other;

Query Match 100.0%; Score 4847; DB 21; Length 4847;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4847; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGCCTCATGCTCCGGAGGCTGTGAGCGGCTGGCGCGTGTCTCTGGAGCAGGGGGCCAG 60
1 ggcctcatgctccggaggctgtgagcggctggcgcggtgtctctggagcagggggccag 60
61 GAATTCGATGTAATACTAACAAGTCTGAGCCCTGGACCTCCACACAGAGATGAA 120
61 gaattcgaatgtaataactaacaagtctgagccctggacccctccacacagagatgaa 120
62 gaattcgaatgtaataactaacaagtctgagccctggacccctccacacagagatgaa 120
121 GGATATGACATAGAGAAAGATATATATCCCAATCTCGGTATAGAAAGTGTAGGGA 180
121 ggaatgacatagagaaagatataatcccaatctcgggtatagaaagtgtaggga 180
122 ggaatgacatagagaaagatataatcccaatctcgggtatagaaagtgtaggga 180
181 GAGAACACAGACTTCTGGAGCGCACAGACCCGTAAGATTCCAAATTCAGAGAACTCG 240
181 gagaacacagacttctggagcgcacagacccgtaagattccaaattcagagaaactcg 240
241 ACCGTTGGAATGCCAATGATGCTTGAAGACAGCCGAGCCGAGGCGCTCTCTTGA 300
241 accgttggaatgccaatgatgcttgaagacagccgagccgagcgctctcttga 300
242 accgttggaatgccaatgatgcttgaagacagccgagccgagcgctctcttga 300
301 TGCCCTCATGATCTCAGCTCAGAAATCTTGATAGAGAGCATCCCAAGGAGAAATACA 360
301 tgccctcatgatctcagctcagaaatcttgatagagagcatcccaaggagaaataca 360
361 TCATGCGTTAGTCTCTGAAAGCCATCCGACACTACTTGGAAAACACAGCCACAGTGA 420
361 tcatgcgttagtctctgaaagccatccgactacttggaaaacacagccacagtgga 420
421 CAATGCGGGGCTTTTTCGATGACCTTTTTCGATGACCTTTTTCGATGACCTTTTTCG 480
421 caatgctgggcttttttcgatgaccttttttcgatgaccttttttcgatgaccttttttcg 480
481 CCACAGAAAGGGAGGAGCTCTCAATGAGAGAGCTGTGTCTCTGTCACAGAGCTTTC 540
481 ccacagaaaggaggagctctcaatgagagagctgtgtctctgtcacaagagcttttc 540
541 TGACGTGAATGCGAAGACACTAGAGACTGTGGCAGAGAGAGCTGAATGAATGGGCC 600
541 tgacgtgaatgcgaaagacactagagactgtggcagagagagctgaatgaatgggcc 600
601 AGAGCGTGCCTCCCTGGAGAGGTTGTGTGATCTTTCGCGCAGCAGAGCTCATCTGTC 660
601 agagcgctgcctccctggagagggtgtgtgatcttttcgcgacagagctcatctgtc 660
661 CATGCTGCTGCTGATGATCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
661 catgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 720
721 ACACCTCTTGAGATATACCCAGGACAGAGCTGAACCTGCAATGACCTTGTGTAGT 780

721 acacctcttgagatataccagcaacagagcttaaccttcagtaacgctgtgttagt 780
781 GCTGGGCTCTCTCTGACGAGAAATCGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 840
781 gctgggctctctctgacgagaaatcgtggctgtgctgctgctgctgctgctgctgct 840
841 GAATTCACGAAACCGGTCTCCGCTTGGCGGGGCATCTTAACATGGCATTTTAAAGAT 900
841 gaattcacgaaacccgtctccgcttggcggggcatttaacatggcatTTTAAAGAT 900
901 CCTTAAGTAAAGAACTTAAAGAGAAATCCCTGGTGGAGCTCATCAATTTGCTCCAA 960
901 ccttaagttaaagaaacttaaagaaatccctgggtggagctcatcaatTTGCTCCAA 960
961 CGATGGGACAGAAATGTTTGAAGCAGACCGCTTGGACGCTGCTGCTGAGAGACCCGT 1020
961 cgatgggacagaaatgTTTGAAGCAGACCGCTTGGACGCTGCTGCTGAGAGACCCGT 1020
1021 TGTTGCCATCTTAGCATGATTTTAAATGTAATTTCTGGGACCAACAGGCTTCGGG 1080
1021 tgttgccatcttagcatgatTTTAAATGTAATTTCTGGGACCAACAGGCTTCGGG 1080
1081 ATCAGCTGTTTATCCCTTTTACCCAGCAATGATGTTTGCATCAGCGCTCACAGCATA 1140
1081 atcagctgtttatcccttttaccagcaatgatgTTTGCATCAGCGCTCACAGCATA 1140
1141 TTTCAGAGAAATGCGTGGCGCCGACGATGAACGTGTCCAGAAATGAATGAATGTTCT 1200
1141 tttcagagaaatgCGTGGCGCCGACGATGAACGTGTCCAGAAATGAATGAATGTTCT 1200
1201 TACTTACATTTAAATTTAAATGTAATGCTGGCTGAAGGCAATTTCTCAGAGTGTCA 1260
1201 tacttacaTTTAAATTTAAATGTAATGCTGGCTGAAGGCAATTTCTCAGAGTGTCA 1260
1261 GAAATCCGCGAGGAGAGAGCGCTCGATATTGGAAGAAAGCCGGTACTTCCAGAGCATCAC 1320
1261 gaaatccgCGAGGAGAGAGCGCTCGATATTGGAAGAAAGCCGGTACTTCCAGAGCATCAC 1320
1321 TGTGGGTGTGGCTCCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
1321 tgtgggtgtggctccatTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
1381 GACCTGGGCTTCATCTGACAGCAGACAGAGCTTTCACAGTGTGACAGCTTCATATC 1440
1381 gacctgggcttcattctgacagcagacagagcttTCACAGTGTGACAGCTTCATATC 1440
1441 CATGACTTGTGCTTGAAGTAACACCGTTTCACTAAGTCCCTTCAGAAAGCTCAGT 1500
1441 catgacttGTGCTTGAAGTAACACCGTTTCACTAAGTCCCTTCAGAAAGCTCAGT 1500
1501 GCGTGTGACAGATTTAAGATTTGTTCTAATGAAAGAGTTTCATATGATTAAGAAACAA 1560
1501 gcgTGTGACAGATTTAAGATTTGTTCTAATGAAAGAGTTTCATATGATTAAGAAACAA 1560
1561 ACAGAGCAGTCTCAGACATCAAGATAGAGATGAAGAAATGCCACCTTGGATGGAGCTCCTC 1620
1561 acagagcagTCTCAGACATCAAGATAGAGATGAAGAAATGCCACCTTGGATGGAGCTCCTC 1620
1621 CCACCTCAGTATCCAGAACTGGCCCAAGCTGACCCCAATTAAGAAAAAGAACAGAGGGC 1680
1621 ccacctcagTATCCAGAACTGGCCCAAGCTGACCCCAATTAAGAAAAAGAACAGAGGGC 1680
1681 TTTCAGGGGCAAGAAAGAGAGTGTAGGACGCTCAGCGCAGCTAGCATGACAGCGGTGCT 1740
1681 ttTCAGGGGCAAGAAAGAGAGTGTAGGACGCTCAGCGCAGCTAGCATGACAGCGGTGCT 1740
1741 GCGAGAGAGAAAGGACCACTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
1741 gCGAGAGAGAAAGGACCACTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
1801 AGAAGGCAAGCAATCCACCTGGGACACTGCGCTTACAGAGGACAGCTGACACATCGA 1860
1801 aGAAGGCAAGCAATCCACCTGGGACACTGCGCTTACAGAGGACAGCTGACACATCGA 1860

Db 1801 agaagcaagacacatccacccctgagccacctgccttaccagagacatgacacagcatcga 1860
 QY 1861 tcttgagatnccagaggggtaaacctggttggaatctgcggcatctggtggagagcaaaaaac 1920
 Db 1861 tcttgagatcccaagagggtaaacctggttggaatctgcggcatctggtggagagcaaaaaac 1920
 QY 1921 cctctcattccagacatttccagacatttccagacatttccagacatttccagacatttccagac 1980
 Db 1921 cctctcattccagacatttccagacatttccagacatttccagacatttccagacatttccagac 1980
 QY 1981 tggaaacctgcgttatgagccagcagcctgcgttatgagccagcctgcgttatgagccagcctgcgt 2040
 Db 1981 tggaaacctgcgttatgagccagcagcctgcgttatgagccagcctgcgttatgagccagcctgcgt 2040
 QY 2041 catctctgttgggaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaat 2100
 Db 2041 catctctgttgggaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaat 2100
 QY 2101 ccttgagccctgac 2160
 Db 2101 ccttgagccctgac 2160
 QY 2161 agccaacctgagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
 Db 2161 agccaacctgagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2220
 QY 2221 cagagacatctac 2280
 Db 2221 cagagacatctac 2280
 QY 2281 catctctgttgggaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaat 2340
 Db 2281 catctctgttgggaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaat 2340
 QY 2341 ccaagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaac 2400
 Db 2341 ccaagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaacagtaac 2400
 QY 2401 ggaag 2460
 Db 2401 ggaag 2460
 QY 2461 taaacctgtgcgag 2520
 Db 2461 taaacctgtgcgag 2520
 QY 2521 tttcacagaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaag 2580
 Db 2521 tttcacagaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaagaaatgatgaaag 2580
 QY 2581 agtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2640
 Db 2581 agtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2640
 QY 2641 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2700
 Db 2641 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2700
 QY 2701 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2760
 Db 2701 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2760
 QY 2761 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2820
 Db 2761 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2820
 QY 2821 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2880
 Db 2821 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2880
 QY 2881 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2940
 Db 2881 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 2940

QY 2941 agcgttccctccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000
 Db 2941 agcgttccctccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3000
 QY 3001 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3060
 Db 3001 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3060
 QY 3061 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3120
 Db 3061 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3120
 QY 3121 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3180
 Db 3121 cttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3180
 QY 3181 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3240
 Db 3181 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3240
 QY 3241 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3300
 Db 3241 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3300
 QY 3301 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3360
 Db 3301 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3360
 QY 3361 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3420
 Db 3361 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3420
 QY 3421 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3480
 Db 3421 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3480
 QY 3481 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3540
 Db 3481 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3540
 QY 3541 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3600
 Db 3541 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3600
 QY 3601 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3660
 Db 3601 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3660
 QY 3661 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3720
 Db 3661 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3720
 QY 3721 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3780
 Db 3721 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3780
 QY 3781 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3840
 Db 3781 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3840
 QY 3841 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3900
 Db 3841 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3900
 QY 3901 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3960
 Db 3901 tttgagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 3960
 QY 3961 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 4020
 Db 3961 ggcagcagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaacagtaaac 4020

[illegible]

| FT | CDS | 116..4429 |
|----------------------------|---|-----------|
| FT | | /tag= a |
| XX | | |
| PN | W09846736-A1. | |
| XX | | |
| PD | 22-OCT-1998. | |
| XX | | |
| PE | 16-APR-1998; 98WO-US07673. | |
| XX | | |
| PR | 16-APR-1997; 97US-0843459. | |
| PA | (MILL-) MILLENNIUM PHARM INC. | |
| PI | Shyjan A; | |
| XX | | |
| DR | WPI: 1998-568724/48. | |
| XX | | |
| PT | P-PSDB: AAM80597. | |
| XX | | |
| PS | New isolated multidrug resistance-associated polypeptide - used to develop products for modulating multidrug resistance, particularly for reducing resistance of tumours to chemotherapeutic drugs | |
| XX | | |
| XX | Claim 1; Page 54-59; 93pp; English. | |
| CC | This nucleotide sequence codes for novel human multidrug resistance associated protein (MRP-beta, see AAM80597), over-expression of which is thought to be associated with the emergence and/or persistence of a multidrug-resistance phenotype in transformed mammalian cells, including carcinoma and adenocarcinoma cells. A unique fragment (see AAM65683) of the MRP-beta gene was identified by computer assisted searching of a nucleic acid database corresponding to a human endothelial cell (HUVVEC) expression library. It was used to screen the HUVVEC expression library. This yielded a 4.78 Kb clone, designated fohnd013a05m (deposited as ATCC 98409). Two independent cDNA clones comprising approximately 60 residues upstream from fohnd013a05m were isolated by hybridisation screening of human brain and liver cDNA libraries using a probe from the 5' end of fohnd01305m. The presented sequence of MRP-beta cDNA comprises fohnd01305m and 66 additional upstream nucleotides. The invention provides compositions and methods for improving the effectiveness of chemotherapeutic regimens to eradicate multidrug-resistant transformed cells from the body of a mammal, especially a human. The disclosed compositions include MRP-beta nucleic acids, including probes and antisense oligonucleotides (see also AAM65684-88), MRP-beta polypeptides and antibodies, MRP-beta expressing host cells, and non-human mammals that are transgenic or nullizygous for MRP-beta. The disclosed methods include methods for attenuating aberrant MRP-beta gene expression, protein production and/or protein function, and for improving the effectiveness of chemotherapy for a mammal afflicted with a multidrug-resistant tumour, wherein the tumour is of mammary, respiratory tract, urogenital tract, endocrine system or immune system origin. In addition, methods are disclosed for identifying and using a modulator, such as an inhibitor, that is cytotoxic to cells expressing MRP-beta. | |
| XX | | |
| SO | Sequence 4847 BP; 1225 A; 1179 C; 1263 G; 1180 T; 0 other; | |
| Query Match | 100.0%; Score 4845.4; DB 19; Length 4847; | |
| Best Local Similarity | 100.0%; Pred. No. 0; | |
| Matches 4846; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; | |
| OY | 1 GGCCTATGCGTGGAGCGGTTGAGCGGCTGGCGGGTTCCTGGAGCAGGGCGCAG 60 | |
| Db | 1 ggcctatgctggggagcgtgtgttgagcggctgcgcggtgtccctggagcagggcgccg 60 | |
| OY | 61 GAATTCGTGATGTAACATCAACAGCTGTGTAGCCCTGGAACCTCCACTCAGAGAAGATGA 120 | |
| Db | 61 gaattctgtagtgaacatacaacacagctctgtgagccctcggaaccccaactcagaagaatgaa 120 | |
| OY | 121 GGATATCGACATAGAAAAGAGTATATATATCCCAAGTCTGGGTATAGAAATGTGAGGGA 180 | |
| | | |

| | | | | |
|---|---|------|--|------|
| D | b | 121 | ggaatcgcacataggaagaagatatacabcaccccaagtcctctggtatagaaagtgtgaaggga | 180 |
| O | y | 181 | gagaaacacagacactcttctggagacgcacacagacacccgtgaagattccaaattccagagaacacgcg | 240 |
| D | b | 181 | gagaaacacagacactctcttggtgacgcacgaagagacgcgtgaagattccaaagtctcaagagaactcgc | 240 |
| O | y | 241 | accgcttggaattgccaaagatgccttgggaaacacagacgccgacccgaagggcccttccttga | 300 |
| D | b | 241 | accgctggaatgcacaaatgctcttggaacacagacgcgcgaagggcccttccttga | 300 |
| O | y | 301 | tgccctccatgcatcttcacgcttcagactccgatcttgatgagacgactcccaagggaaattacca | 360 |
| D | b | 301 | tgccctccatgcatcttcacgactccgaatccctgtgtgagagacatcccaagggaattacca | 360 |
| O | y | 361 | tcagagcgttgagtgctctgtagaagcccatccgaatcttgcaaacaccgcaccccaagtgga | 420 |
| D | b | 361 | tcagagcgttgagtgctctgtagaagcccatccgaatcttgcaaacaccgcaccccaagtgga | 420 |
| O | y | 421 | caattgcctggcctttttttccgtatagacttttttcgtggcctttctcttgcccgctgtggc | 480 |
| D | b | 421 | caatgcctgggcctttttttccgtatagacttttttcgtggcctttctcttgcccgctgtggc | 480 |
| O | y | 481 | ccacaaagaaaggagagctcccaaggaaagcgtgtgctctctgtccaaagcagagcttc | 540 |
| D | b | 481 | ccacaaagaaaggagagctcccaaggaaagcgtgtgctctctgtccaaagcagagcttc | 540 |
| O | y | 541 | tgacgctgaaactgcagagacttagagacactgttgcaagaagacgtgaatgaagttggcc | 600 |
| D | b | 541 | tgacgctgaaactgcagaaagacttagagagacactgtgtgcaagaagacgtgaatgaagttggcc | 600 |
| O | y | 601 | agacgctgtccttcctgcgaagggtgtgtgtgatcttctgcgcacacagcctcatcctgtc | 660 |
| D | b | 601 | agaagcgtgtccttcctgcgaagggtgtgtgtgatcttctgcgcacacagcctcatcctgtc | 660 |
| O | y | 661 | catgctgtgacctgaaatgacacgacgctggccttcacatggaacacacgcttcacatggttaa | 720 |
| D | b | 661 | catgctgtgacctgaaatgacacgacgctggccttcacatggaacacacgcttcacatggttaa | 720 |
| O | y | 721 | acacctcttgagatatacccaagcaaaaggtctaaacctgcgaataagcttgttgtagt | 780 |
| D | b | 721 | acacctcttgagatatacccaagcaaaaggtctaaacctgcgaataagcttgttgtagt | 780 |
| O | y | 781 | gctgggcctcctcctgtacgaaatcgtgcgtctctgtgccttgacttgagcacttggcatt | 840 |
| D | b | 781 | gctgggcctcctcctgtacgaaatcgtgcgtctctgtgccttgacttgagcacttggcatt | 840 |
| O | y | 841 | gaaattacggaacgsgtgcgcttggggggggcacatccmaacatngacatttaaaagaat | 900 |
| D | b | 841 | gaattacacgaacccggtgtccgcttggtgggggggcacatcccaacatgagcatttaaaagaat | 900 |
| O | y | 901 | cctttaagtttaagaacattttaagaagaattccctgcggtgagctcatcaaatltgtccaa | 960 |
| D | b | 901 | ccttaagtttaagaacattttaagaagaattccctgcggtgagctcatcaaatltgtccaa | 960 |
| O | y | 961 | cgatgaggaagaaatgttttagaggaagacgcttgacgctgtgctgaggaacgcgt | 1020 |
| D | b | 961 | cgatgaggaagaaatgttttagaggaagacgcttgacgctgtgctgaggaacgcgt | 1020 |
| O | y | 1021 | tggtggccattcttagcattgatttatattgataattattcttgagacacaaagccttccctgg | 1080 |
| D | b | 1021 | tggtggccattcttagcattgatttatattgataattattcttgagacacaaagccttccctgg | 1080 |
| O | y | 1081 | atcagcgtgttttcttctcttcttcaaccagaaatgatatgtttgcatcagcgctcaacgcata | 1140 |
| D | b | 1081 | atcagcgtgttttcttctcttcttcaaccagaaatgatatgttttgcacatcagcgctcaacgcata | 1140 |
| O | y | 1141 | tttcaggaagaaatgcgtggccgcgcacagatgaacgctgtccaaagaatgaatgaattct | 1200 |
| D | b | 1141 | tttcaggaagaaatgcgtggccgcgcacagatgaacgctgtccaaagaatgaatgaattct | 1200 |
| O | y | 1201 | tactacacattaaatttcaaaatgataatgctgtgtgcaagaacatttcccaagagtcca | 1260 |
| D | b | 1201 | tactacacattaaatttcaaaatgataatgctgtgtgcaagaacatttcccaagagtcca | 1260 |

| | | | |
|----|------|--|------|
| QY | 1261 | GAAATTCGCGAGGAGGAGGCGTCGAGATTGTGAAAAAGCCGGGTACTTCCAGAGCATCAC | 1320 |
| DB | 1261 | gaaatctcgagaggaagagcgctcgatattggaaaaagccgggtacttccagagatcac | 1320 |
| QY | 1321 | TGTGGGTGTGGCTCCCATTTGTGTGTGATTTGCCAGCGTGTACCTTCTCTGTTCATAT | 1380 |
| DB | 1321 | tgtggtgtgtgctcccatctgtgtgtatcttgcaacgctgtgtacctctctgttcatat | 1380 |
| QY | 1381 | GACCCGTGGGCTTGCATCTGACAGAGACAGGGCTTTCACAGNGTGTACAGTTCATATTC | 1440 |
| DB | 1381 | gaccttggtcttgatctcgaaagaagacagcttctcaagtggtgcagcttctcaatlc | 1440 |
| QY | 1441 | CATGACTTTTGTCTTGAAGTAACACCGTTTTCAGTAAAGTCCCTTCAGAAAGCCCTCAGT | 1500 |
| DB | 1441 | catgacttttgcttgaaagtaaacacccggttttcagttaaagttccctctcaagaagctcagt | 1500 |
| QY | 1501 | GGCTGTTCACGATTTTAAGAGTTTGTTCATATGGAAGAGTTCCACATATTAACACAA | 1560 |
| DB | 1501 | ggctgttgcagatttaagaattgtgttctcaatggaagagttctacaatgataaagaaca | 1560 |
| QY | 1561 | ACCACACGCTCTCAACATCAAGATAGAGATGAAGAAATGACACCTTGGATGGACCTCTC | 1620 |
| DB | 1561 | accacacgctctccatcacatcaagatagagaatgaanaatgcacctgtgcctgtgactctc | 1620 |
| QY | 1621 | CCATCCACGATATCCAGAACTCGCCCAAGCTGACACCCCAAAATGAAAAAGACAAAGAGGC | 1680 |
| DB | 1621 | ccatccacgattatccagaactcgcccaagctgcaccccaaatgaaaaaagacaagaaggcc | 1680 |
| QY | 1681 | TTCCAGGGGCGAAAGAAAGAGGTGAGGGCAGCTGCAGCGACTGAGCATTCAGCGGTGCT | 1740 |
| DB | 1681 | ttccagggcgaaagaaagaggtgagggcagctgcagcgactgcagcatcacaagcggtgct | 1740 |
| QY | 1741 | GGCAGAGGAGAAAGGCCACCCCTCCCTGGACATGACGAGCGGCCCAAGTCCGGAAGAGA | 1800 |
| DB | 1741 | ggcagagagaaagggcacccctccctggacatgacgagcgcccaagtgccgaagagga | 1800 |
| QY | 1801 | AGAAAGCAGAGCAGATCCACCTGGGGCCACACGCGCTTACAGAGGACACTGCACAGCATCGA | 1860 |
| DB | 1801 | agaagcgacagacatccaccctgggccacctgcgcttaacagagacaactgcacagatcga | 1860 |
| QY | 1861 | TCTGGAGATCCAAAGGGTAAACCTGGTTGGAAATCTCGGGCAGTGTGGAAATGGGAAAAAC | 1920 |
| DB | 1861 | tctggagatccaaagggtaaacctgggttggaatctcgggcagtgtggaaatgggaaaaac | 1920 |
| QY | 1921 | CTCTCTCATTTTCAGCAATTTTAAAGGCCAGATGAGAGCGCTTACAGAGGAGGATTCGACATCG | 1980 |
| DB | 1921 | ctctctcatlctcagcatltaaggccagatgaagcctctcagaagggagcatgtgacatcg | 1980 |
| QY | 1981 | TGGAACTTTCGCTTATGTGTGGCCAGCAGGCGCTGATCTCAATGATCACTCTGAGACAA | 2040 |
| DB | 1981 | tggaaacttctgcttatgtgtggccagcagcgctgatatctcaatgactactctgagagcaa | 2040 |
| QY | 2041 | CATCTCTGTTTGGGAAGCAATATGATGAAGAAAGATACACTCTGTCTCTAACAAGCTGCTG | 2100 |
| DB | 2041 | catctctgtttgggaagcaatatgataagaagaatacaactctgtctgaacaagctgtctg | 2100 |
| QY | 2101 | CGTGAAGGCTTACCGTGGCAATTTCTCCACACAGCAAGCTTACGCGAATTTGGAGAGCGAGG | 2160 |
| DB | 2101 | cgtgaaggcttaccgctggcaatttctccacacagcaagcttgaacggaatttggaagcgagag | 2160 |
| QY | 2161 | AGCCAACTTGAGCGGTGGGCGAGCCAGAGAGATCAGGCTTGGCCCGGCTTATATAGTA | 2220 |
| DB | 2161 | agccaaacttgagcggtggcgagccagagagatcagagcttggcccggtctgtataatgta | 2220 |
| QY | 2221 | CAGGAGCATCTACATCTGTGACGACCCCTCAGTGCCCTTAGATGCCATGTGGACAACA | 2280 |
| DB | 2221 | caggagcatctacatctgtgcgacccctcagtgcccttagatgcccacatgtggacaaca | 2280 |
| QY | 2281 | CATCTTCAATATAGTGTATCCGGGAAACATCTCAAGTCCAGACAGTTCTGTTTGTATACCA | 2340 |
| DB | 2281 | catcttcaatattagtgatccggaaacatctcaagtcacaaagcagttctgtttgttacc | 2340 |

2341 CCAGTACAGTACCTGGTGTGACTGTGATGAAGTATCTTCAATGAAAGAGGGCTGTATTAC 2400
2241 ccagttacagtaacctggtgtgactgtgataagtgatctcttcataagaagggctgtattac 2400
2401 GGAAGAAGGCAACCATATGAGAACATGATGAATTTAAATGGTGACATATGCTACCATTTTAA 2460
2401 ggaagaagagccacatagaagaaactgataatcttaaatggtactatgtactacattttaa 2460
2461 TAACTTTGCTGGAGAGAGACACCGCAGTGTGAGATCAATCAAAAAAGAAACAGATGG 2520
2461 taacctgttctgtggaagagacacccgcaagttgagatacattcaaaaaagaaacagttgg 2520
2521 TTCACAGAGAAGTCCACAGACAAAGGCTCTTAAACAGAGATCAATTAAGAAGAAAAAGC 2580
2521 ttccagaaagaagatcacaaagaagggctccataaacaagatcaataaagaagaaaaagc 2580
2581 AGTAAACCCAGAGAAGGAGCGCTTGTGCAAGCTGGAAGAAAAGGCAAGGTTTCAGTCC 2640
2581 agttaaagccagaagaagagcagctgtgcagctggaagagaagagcagggltcagtgcc 2640
2641 CTGCTCAGTATATGTTGTCATCCACAGGCTGTGGGGCCCTTGGCATTCCTGTTAT 2700
2641 ctggtctcagatatagtgtgtctacacacccgctgctggtggcccttgscattccgtttat 2700
2701 TATGGCCCTTTTCAATGCTGAATGTAGCAGCACCCGCTTCAGCACTGTGGTTAGTTA 2760
2701 tatggcccttttcatgtctgaatgtagaagcaacgccttcacagcctgtgtgtgagtta 2760
2761 CTGATACAGAGAGAGCGGGAACCACTGTCAGTGCAGGAGAAAGCAAGCTGGGTAG 2820
2761 ctgagtaacaagaagaagcgggaacacacgtagctcgaaaggaaacgagacctggtag 2820
2821 TGAGAGCATGAGAGACATCTCATATGACATATATGACAGATCTACGCTCTCCAT 2880
2821 tgaagcatgaagagaacaactcctcatatgcagtaactatgccaagcatctaacgcccctcat 2880
2881 GGCAGTATGCTGATCCTGGAAGCCATTCCAGAGATGTCTTTTGTCAAGGCAAGCTGCG 2940
2881 ggcagtatgtctgactcttgaaagcattcgaagagttgtcttctcgaaggcacgctgag 2940
2941 AGCTTCCTCCGGCTGCATGACGAGCTTTTCCGAAGATCCTTGAAGCCCTATGAAGTT 3000
2941 agcttctcccggtctgcatgaagcagcttctccgaagatctctcgaagccctatgaagtt 3000
3001 TTTTGACACGACCCCAACAGGAGATTTCCACAGGTTTTCCAAAACATGATGAAGT 3060
3001 ttttgacacgacccccaagagagatcttcaacagtttccaaaacatgagttagt 3060
3061 TGAGTGCAGGCTGCAGTTCCAGGCGGAGATGTATCATCAAGAACTTATCCTGGTCTT 3120
3061 tgaagtgcaggtgcagttccaggtccgagatgttcatccagaagcttactcgtgtctct 3120
3121 CTGTGTGGGATGATCCGACAGAGTCTTCCCGTGGTTCCTTGTGACAGTGGGCCCTTGT 3180
3121 ctgtgtgggatatgattccgacagagcttcccggtgttctctgtgagatggggcccttgt 3180
3181 CATCTCTTTTCACTCTGACATTTGTCTCCAGGGTCTTGATTTGGGAGCTGAAGCTCT 3240
3181 catctcttttcactctgacattgtctccaaggtctcgtgttcggagctgaaagcgtct 3240
3241 GGAACAATACAGGAGTACACTTTCCTCTCCACATCATCAGTCCAGCATACAGGGCTTGC 3300
3241 ggaacaatatcagagagtaaccttctctccacatacagctccagcatalaagggccttgc 3300
3301 CACCATCAGCGCTACATATAAGGAGAGATTTCTGCACAGATACAGAGAGCTGTGA 3360
3301 caccatccacgctcaacataaagaagagttctctgacacagataacacagagctcgtgga 3360
3361 TGACAAACAAAGCTCTTTTGTGTTTACGTGTGCAATGCGGTGCTGTGGCT 3420
3361 tgacaaacaaagctcttcttcttcttgaagtgtagatgagtggtggtggtggtggt 3420
3421 GGACCTATACAGCATGCCCTCATACACACAGGGGCTGATGATGTTTATGACAGG 3480

3421 ggaacttaacagcatcgcctcccatcacacacacagggctgtgactgtcttcttgcacgg 3480
3481 GCAGATTCCTCCAGCGATATGGGGTCTGCCATCTTATCTCTGTCGCAAGTAAAGGAGCT 3540
3481 gcaatctcccacagcctatggtggtctccgcatctcttatagtgtctcagttaaaggggt 3540
3541 GTCCAGTTTACGGTCAAGTACGATCGCATGAGACAGAAAGCTTGATTCACCTGGTGAAG 3600
3541 gtccagtttaacggttcaagactcgtgcatctgagacagaagctcgtatccactcgtggaag 3600
3601 GATCAATCACAATTAAGATCTGTCTTGGAGACACTCTCCAGAAATTAAGAACAAAGC 3660
3601 gatcaatcacatactaagactcgtctcttgaagcacctgtccagaatlaagaacaagc 3660
3661 TCCCTCCCTCAGTACGGCCCGAGAGGAGAGTGAACCTTTCAGAACGACAGATGAGTA 3720
3661 tccctcccctgactggtcccaggaaggaaggtgaactcttgagaacagagatagaagta 3720
3721 CCGAAGAAAACCTCCCTCTGCTCTTAAAGAAATATCTTACGATCAAACTTAAAGAA 3780
3721 ccgaagaaacctccctctcgtcttaagaagtaaccttcaacgatacaaaacttaagaagaa 3780
3781 GATTGGCAATTTGGGGCGGACAGATCAGGGAAGTCTCGCTGGGATGGCCCTTCCG 3840
3781 gattggcaatttggggcggaacagatcagggaaagtcctgctgggattggcccttccg 3840
3841 TCTGTGAGAGTTATCTGAGAGCTGCATCAAGATGATGAGATGAGATCAATGATATTGG 3900
3841 tctgtgagagttatctgagagctgcatacaagatgtagatgagatgagatgattgg 3900
3901 CTTTGCACACTCCGAAGCAAACTCTATCATTTCTTCAAGAGCCGGTGTTCAGTGG 3960
3901 ctttgcacctccgaagcaaaactctatcatcttccataagaagccggtggttccagtg 3960
3961 CACTGTCAGATCAAAATTTGGACCCCTTCAACAGTACACGAAACAGATTGGAGTGC 4020
3961 cactgtcagatcaaaattggaccccttcaacagtaacgtgaagacagatcttggaagc 4020
4021 CTTGGAGAGACACATGAAAGATGATTTGCTCAGTACCTTGAAACTTGAATCTGA 4080
4021 cttggagagacacatgaaagatgattgtcagctacactctgaaacttgaatctga 4080
4081 AGTATGAGAAATGGGATTAATCTTCACTAGTGGGGAACGGCAGCTTGTGTACATGACTAG 4140
4081 agtatgagaatgggataatcttcaagttcagtggtggaacgagatctgtgcatagtag 4140
4141 AGCCCTGCTCCGCCACGTGTAAGATTTGATTTAGATGAACCAACAGCTGCATGAGAAC 4200
4141 agccctgctccgccacgtgtaagatttgatttagatgaagccaagctgcataagagac 4200
4201 AGAGACAGACTTATTTGATTTCAAGAGACCATCCGAAAGCATTTTCAGACTGTACATGCT 4260
4201 agagacagacttattgtattcaagaagacatccgaaagcatttcagacgtgtacatgct 4260
4261 GACCATTTGCCATCGCTGACACAGGTTTACGCTCCGATAGATTTATGCTGGGCCCA 4320
4261 gaccatttgccatcgctgacacaggttctaggtctcgatagatltatgtggtcgtgcca 4320
4321 GGGACAGGTGAGGATTTGACACCCCATCGTCTCTGTCACAGCAACATTTCCCATTT 4380
4321 gggacaggtgaggtttgacaccccatcgctctcttcgtccacaagatctccagatt 4380
4381 CTATGCCATGTTTGTCTGTGACAGAACAAAGTGTGTCAGAGGCTGACTCTCCGTGT 4440
4381 ctatgccaattgttctgtctgtagagaagaagtgctgttcaaggtcgtacccctcgt 4440
4441 TGACGAAGTCTCTTTTCTTGAAGCATTTGCCATTCCTGCTGGGGGCGGCCCTTATC 4500
4441 tgacgaagtctcttttcttgaagcatgtccatccctggtggtggtggtggttctc 4500
4501 GCGTCCTCTACGAAACCTTGCCTTCTGATTTTATCTTGCACAGAGCTTCGGAT 4560

| | | | | |
|----------|--|--|----|------|
| Db | 4501 | gcgcctccctccacgaaacctgacctctctcctcgatttctcttcgcacagcagttccg | at | 4560 |
| Qy | 4561 | TGGCTTGTCGTCATCTTTTATGAGGAGATCATATTTTGTATTTATTTTTCATAT | | 4620 |
| Db | 4561 | tgctcttgcttgcttccactttttagggagagtcatacttttgatttgcatttccat | | 4620 |
| Qy | 4621 | TCATGTAAACAAATTTAGTTTTGTTCTTAATTTGCACCTCTAAAAGGTTACGGACCGT | | 4680 |
| Db | 4621 | tcaagtacaacaatttagtttcttgccttaattgcactcaaaagttcaaggaaacgc | | 4680 |
| Qy | 4661 | TATATATATGTCATCAGGCGCTTAATGAAAGCTTTAGGTGAGCTATATCTATAT | | 4740 |
| Db | 4661 | tattataattgacacagagccctatacgaagcttatacgtgtaagctatactcata | | 4740 |
| Qy | 4741 | AATCTGTACATACCCATATTTTACATCGAAAAATGTAAGCTGTTATTTATTAATAAT | | 4800 |
| Db | 4741 | aattctgtacctagcctatactattccagtgaaatagtaagctglttattctataat | | 4800 |
| Qy | 4801 | AAGCATGTGCTAAAAAAGGCGGCGCGC | | 4847 |
| Db | 4801 | aagcactgtgctaaaaaagcgcgcgcg | | 4847 |
| RESULT 3 | | | | |
| ID | AAZ30079 | standard; cDNA; 5838 BP. | | |
| XX | AAZ30079; | | | |
| XX | AAZ30079; | | | |
| DT | 26-JAN-2000 | (first entry) | | |
| XX | | | | |
| DE | | cDNA encoding a human MPR-related ABC transporter designated MOAT-C. | | |
| XX | | | | |
| KM | | Human; MPR-related ABC transporter; MOAT protein; MOAT-C; | | |
| KW | | MOAT mediated transport; anticancer drug sensitivity; | | |
| KW | | transporter mediated cellular efflux; anticancer; ss. | | |
| XX | | | | |
| OS | | Homo sapiens. | | |
| XX | | | | |
| FH | Key | Location/Qualifiers | | |
| FT | CDS | 126..4439 | | |
| FT | | /*tag= a | | |
| FT | | /product= "MOAT-C" | | |
| FT | | /note= "MPR-related ABC transporter" | | |
| XX | | | | |
| PN | W09949735-A1. | | | |
| XX | | | | |
| PD | 07-OCT-1999. | | | |
| XX | | | | |
| PF | 26-MAR-1999; | 99MO-US06644. | | |
| XX | | | | |
| PR | 27-MAR-1998; | 98US-0079759. | | |
| PR | 03-AUG-1998; | 98US-0095153. | | |
| XX | | | | |
| PA | (FOX-C-) FOX CHASE CANCER CENT. | | | |
| XX | | | | |
| PI | Kruh G, Lee K, Belinsky M, Bain L; | | | |
| XX | | | | |
| DR | WPI: 1999-610812/52. | | | |
| XX | P-PSDB; MAY43542. | | | |
| XX | | | | |
| PT | New transporter gene useful for screening for anti-cancer drugs | | | |
| XX | | | | |
| PS | Claim 11; Page 134-135; 153pp; English. | | | |
| CC | | | | |
| CC | The present sequence encodes a human MPR-related ABC transporter (MOAT) | | | |
| CC | protein, designated MOAT-C. The protein comprises a multi-domain | | | |
| CC | structure including a tandem repeat of nucleotide binding folds | | | |
| CC | appended C-terminal to a hydrophobic domain, having Walker A and B ATP | | | |
| CC | binding sites and several potential membrane spanning domains. The MOAT | | | |
| CC | nucleic acids are useful for screening a test compound for inhibition of | | | |
| CC | MOAT mediated transport, indicated by restoration of anticancer drug | | | |
| CC | sensitivity, which in turn causes a reduction of transporter mediated | | | |

CC cellular efflux of anticancer agents. MOAT DNA or RNA may be used as
CC probes to detect the presence or expression of genes encoding MOAT
CC proteins. Anti-MOAT antibodies are useful for detecting and quantitating
CC MOAT proteins.

5838 BP; 1422 A; 1432 C; 1509 G; 1475 T; 0 other;

| | | | | |
|-----------------------|--------------|--------------|--------------|-------------|
| Query Match | 98.9% | Score 4792.6 | DB 20 | Length 5838 |
| Best Local Similarity | 99.8% | Pred. No. 0 | | |
| Matches 4809 | Conservative | 0 | Mismatches 9 | Indels 1 |
| | | | | Gaps 1 |

| | | | |
|----|-----|---|-----|
| QY | 1 | GGCTATCTCTGGGAGCGTGGTTAGCGCGCTGGCGGTCTCTGTGAGACGAGGGCGCAG | 60 |
| Db | 11 | ggctaatctctgggagcgctgtgtgagcgctcgcgcgtgtgtctctgtgagcaaggcgca | 70 |
| QY | 61 | GAATTCTATGTGAAACATACAGTCTGTGAGACCCCTGGAACTCCACTCAAGAAATGAA | 120 |
| Db | 71 | gaattctaatgtgaactaacaagctctgaaacctgtgaacctcgcgcagagaagtatgaa | 130 |
| QY | 121 | GGATATCGACATAGAAAAAGAGTATATCATCCCAAGTCTGGGTATGGAAGTGAAGGA | 180 |
| Db | 131 | ggatatcgacatagaaaaagagtatatcatatcccaagctcctgggtatagaagtgtgaagg | 190 |
| QY | 181 | GAGAACCGACACTTCTGGGACACGCACAGAGACCTGGAAGATTCCAACTTCAGAGAACTCG | 240 |
| Db | 191 | gagaacccagcaactctctggagacgcacaagaccgtgaagattccaagttcagagaactcg | 250 |
| QY | 241 | ACCGTTGGAAATGCCAAGTATGCTTGGAAACAGCGCCGAGCCGAGGCGCTCTCTTTGA | 300 |
| Db | 251 | accgtctggaatgccaagatgcctctggaacacgcagccgcagcgagcgctctctcttga | 310 |
| QY | 301 | TGCGTCCATGCAATTCYACGTCAGAAATCCTGGATGAGAGAGCATCCCAAGGAAAAATACA | 360 |
| Db | 311 | tgcctccatgcattctcagctcagcaatctctgattgagagcatctccaaggaagtacaa | 370 |
| QY | 361 | TCATATGCTTAGTGCTCTGAAGCCCATCCGCACTACTTGCACAAACACGACCCACGTGGA | 420 |
| Db | 371 | tcatatgcttagtgctctgaagcccatccgcactacttccaacaacacgacaccagttga | 430 |
| QY | 421 | CAATCTGCGGCTTTTCTCTGATAGACTTTTCTGGCTTTCTCTCTGCGCGGTGAGC | 480 |
| Db | 431 | caatctgcgctcttctctctgtatgaacttctcgttctctctctctctcgtcccgctgagc | 490 |
| QY | 481 | CCACAAGAGGGGGAGCTCTCATGTGAAGAGCTGTGCTCTGTCCAAAGCAGCAGTCTTC | 540 |
| Db | 491 | ccaacaagaggggagcctctcaatggaagagctgtgtctctctgtccaagaagcagctcttc | 550 |
| QY | 541 | TGACGTGAATCTGCAGAGACTAGAGAGACTGTGGCAAGAAAGCTGAATGAAGTTGGCC | 600 |
| Db | 551 | tgaagtgaaactgcagaagactagagagctgtgtgcaagaagagctgaaatgaactgtggcc | 610 |
| QY | 601 | AGACCTCTTCCCTCGCAAGAGGTTGTGTGATCTTCTGCGCACAGGGCTACATCTGTC | 660 |
| Db | 611 | agaagctgtcttccctcgcgaaaggctgtgtgatactctctgcgcgaacagctcatctcgctc | 670 |
| QY | 661 | CATCTGTGCGCTGATGATCAGCAGCACTGGCTGCTTCAAGTGAACACACTTCATAGTGAA | 720 |
| Db | 671 | catctgtgcctgatatgatacaagcagctcgtcgtcttaagtggaacacgctcatcatgtgaa | 730 |
| QY | 721 | ACACCTCTTGGAGTATACCAAGGAAACAGAGCTTAACCTGCAGTACAGCTTGTGTAGT | 780 |
| Db | 731 | acaactcttgagataaccagcaagcaagagcttaactctgagtaacagctgtgttagt | 790 |
| QY | 781 | GCTGGGCTCTCTCGACGAGAAATCGCGGCTTGGTCCGTTCCACTGACTGGAGCATT | 840 |
| Db | 791 | gctgggctctctcctcgcgaagaaatctgtcgtctctgtgccttgacatgacttgggcatc | 850 |
| QY | 841 | GAATTACCGAACCGGTCTCCCTTGGCGGGGCGCATCTTAACATTCGCAATTTAAGAAAT | 900 |
| Db | 851 | gaattaccggaacggtgtcgcgctctgcggggggccatctcaacatgagcatltaagaagat | 910 |
| QY | 901 | CTTTAAGTTAAGACATTTAAGAGAAATCCCTGGGTGAGGCTACTCAACTTGGCTTCAA | 960 |

| | | | |
|---|------|---|------|
| D | 911 | ccctaaagtaagaacatcttaagaagaatccctgggtgagctccatcaacattgctccaa | 970 |
| Q | 961 | CGATGGGCGACAGATGTTTGAAGCGACACAGCCGTTGGACACCTGCTGGGTGAGACCCGT | 1020 |
| D | 971 | cgatgggcagagaatgctttgaagcagcagccgttgcagctcgctgctgtaggaaccgt | 1030 |
| Q | 1021 | TGTTGCCATCTTAGCGATGATTTATATGTAATTAATTTCTGGGACCAACAGCCTTCTGGG | 1080 |
| D | 1031 | tggtgcacctcttagcagatgatttataatgtaattatcttcggaaccaaaagcttcctgg | 1090 |
| Q | 1081 | ATCAGCGTGTATATCCTCTTTTACCACGAATGATGTTTGGATCAGCGCTCACAGCAT | 1140 |
| D | 1091 | atcagcgcttcttaccctcttaccacgaagaatgctttagatcaagcgtctcaagcata | 1150 |
| Q | 1141 | TTTACAGAGAAAATGCTGTGCCCCGACCGATGACGTCGTGCACGAAGATGAATGAATGCT | 1200 |
| D | 1151 | tttaagagaagaatgctgtgycgcgcacagatgaagctgtccagaagaatgatagaattct | 1210 |
| Q | 1201 | TACTTACATTAATTAATATCAAAATGTATGCGTGGGTCAAGAATTTTCTCAGATGTTCA | 1260 |
| D | 1211 | tacttaccattaaatttataccaatgatgatgcctggtgccaagaatcttccagagytcca | 1270 |
| Q | 1261 | GAAATATCCGGAGAGAGAGCGCTGGATATTGGAAAAAGCCGGGTACTTCACAGCATATAC | 1320 |
| D | 1271 | aaaaatccgcagagagagagcgtcgcgatatggaaaaagccggttacttccagtgataac | 1330 |
| Q | 1321 | TGTGGGTCGTGGCTCCCAATTTGGTGGGATGATGTCACCGGTCGAGACTTCTCTGTTCAAT | 1380 |
| D | 1331 | tgtvggtgtgctccctctgtgtgtgtgtatgtgcagcgtgtgtgacacctctctgttccatac | 1390 |
| Q | 1381 | GACCCGTGGGCTTCGATCTGCACAGCACAGCGCTTTCACAGTGTGACAGTCTTCAATTTC | 1440 |
| D | 1391 | gacctgtggtctgcgactgcagcagcaagsgcttcacagtggtgtgacaagctcttcaattc | 1450 |
| Q | 1441 | CATACATTTTGTCTTGAAGTAAACACCCGTTTTCAGTAAAGTCCCTCTCAGAAGCCTCAGT | 1500 |
| D | 1451 | catgactttgtccttgaagaataaacgcgttcttaagaagtcctctccagaagccctcagt | 1510 |
| Q | 1501 | GGCGTTGACAGATTTTAAAGTTTGTGTTCTTAATAGTAAAGAGTTCACATGATTTAAACA | 1560 |
| D | 1511 | ggcgtgtgacagatcttaagaggttgttcttaatgtaagaaggttccacaatgaataaaca | 1570 |
| Q | 1561 | ACGACGACGTCTCTACATCAAGATAGAGATGAAAAATGCCACTTGGCATGGGACTCCTC | 1620 |
| D | 1571 | acgacgacgtctctacatcaatcaagatagatgtaaaaaatgcacacttggatgagactctc | 1630 |
| Q | 1621 | CCAATCCAGTATCCAGACATCTGCCCAAGCTGACCCCAATATGAAAAAAGACAGAGGCG | 1680 |
| D | 1631 | ccactccagatctccagaacatcgcccaagctgacgcccccaaatgtaaaaaagaagaagtc | 1690 |
| Q | 1681 | TTTCCAGGGGCAAGAAAGAAAGAGTGTAGAGCGATGCGACGCTCATGAGCATCAAGCGCTCT | 1740 |
| D | 1691 | ttccaggggcagaagaagaagaagtgtagcgtgcagcgtgcactgtgcatctcagggcggtct | 1750 |
| Q | 1741 | GGCAGACAGAAAAGGCCACTCCTCTCTGTGACAGTGAAGCGGCCCCAGTCCGAAAGGA | 1800 |
| D | 1751 | ggcagagcagaagaagccactctctctctgtgacagtgtgcagcgccacgttccggaagaga | 1810 |
| Q | 1801 | AGAAGCGAAGCATCTCACCCTGGGCGCACTGGCGCTTACAGAGAGACATGCAACGATATGA | 1860 |
| D | 1811 | agaaggaagaacatccacacttggccacactgcttacaagaagaactgcaacagatctga | 1870 |
| Q | 1861 | TCTGGAATCTCAAGAGAGGTAAATCGTTGGTAATCTCGGCACTGTGGGAAGTGGAAAAAC | 1920 |
| D | 1871 | tcgtgagatccaagaagggtaaacctgcttggaaatctgctgagcgtgtgtggaagtgtgaaaac | 1930 |
| Q | 1921 | CTCTCTCATTTCAAGCCATTTTAAAGCCAGATGACGCTTCTAAGAGGGCAGCATTTGCATCAG | 1980 |
| D | 1931 | ctctctcatcttcagccaattttaggcagagttagcgtctttaggggcagcatgtgcaatcag | 1990 |
| Q | 1981 | TGGAACTTGTGGCTTATGTGGCCACGACAGGCGTGGATCCTCAATGCTACTCTGAGAGCAA | 2040 |

| | | | |
|---|------|--|------|
| D | 1991 | tgnaaccttcgcttatatgctgcccagaagcccgtagctcccaatgctactctcgagagaa | 2059 |
| Q | 2041 | CATCCTCTTTGGGAAGATATGATGAGAAGATACAACTCTGTGTCACACCTGCTG | 2100 |
| D | 2051 | catccgtcttcgggaagaaataatgatagaagaatatcaactctgctcgaaacagctgctg | 2110 |
| Q | 2101 | CCTAAGGCTGACCTGGCCATTCTTTCCCAAGACACCTGACAGGAGATTGGAGAGCCAGG | 2160 |
| D | 2111 | ccctgaagccctgaacctggtccattctcccaagcgacgtcgagagatcttggagcgag | 2170 |
| Q | 2161 | AGCCAACTTGAAGCGGTGGGAGCGCCAGAGAGATACAGCTTGGCCGGGCTGTATAGTA | 2220 |
| D | 2171 | agccaaacctgaagcggtggtggaagcgccaagagatcagcccttcgcygacctgtatagta | 2230 |
| Q | 2221 | CAGGAGCATCTACATCCCTGGACGACCCCTCACTAGTGCCTTAGTGCCATGTGGCAACA | 2280 |
| D | 2231 | caggagatctatactctcttgtagcagccccctcagctgcttagtgccatgtggaacaa | 2290 |
| Q | 2281 | CATCTTCAATATGTGCTATCCGGAAACATCTCAAGTCCAAAGACAGTTCTGTATTACCA | 2340 |
| D | 2291 | catctcaaatagtgctatccggaacaatcctaagtcacaagaacagctctgtgtatccca | 2350 |
| Q | 2341 | CCAATTACAGAACTGGTTTACGTGTATGTAAGGATCTTCATGAAGAAGGGGTGATATAC | 2400 |
| D | 2351 | ccagttacagtaacctggtgtacgtctgtatgaatgatacttcaagaagsgtgcgtatcac | 2410 |
| Q | 2401 | GGAAGAAGGACCCATATAGACATGTGATGTAATTAAATGGTGCATGTACCATTTTAA | 2460 |
| D | 2411 | ggaaagsggcccacatgaggaactgtagtaattaaatggtgctatgctacatctttaa | 2470 |
| Q | 2461 | TAACTCTTGCTGGGAGAGACACCGCCAGTTGATCAATTCAAAAAAGAAACCATGTG | 2520 |
| D | 2471 | taacctgtgctggagagacacgcgcagctgtagatcaatcctcaaaaaggaaccagctg | 2530 |
| Q | 2521 | TTTCCACAAAGACGCACAAACAGAGGGTCTTAAACAGATCAATAAAGAAAGAAAAAGC | 2580 |
| D | 2531 | ttccagaagaagaagccacaagaacaaggtccctaaacaagaatcaagttaaagaagaanaagc | 2590 |
| Q | 2581 | AGTAAACCCAGAGGAAGGAGCGACGCTTGTGCAGCTGTAAGAGAAAGGAGGCTTCAGTGCC | 2640 |
| D | 2591 | agtaaaagccagagaagsgagctctgtgaaagctcggaagsgaaagsgcaggtgtcaagtcc | 2650 |
| Q | 2641 | CTGGTCAGTATATGTGTCTACATCCAAGCTGCTGGGGGCCCTTGGCATTCGTGTGTTAT | 2700 |
| D | 2651 | ctgtgcagtatatgctgtctacataccaagcctgtcggtggcccttgcgtatccctgttat | 2710 |
| Q | 2701 | TATGGCCCTTTTATGCTGTAATGTAGGACGACCGGCTTCACACCTGTGTGGTTACGTTA | 2760 |
| D | 2711 | tatggcccttctcaatgctgtaagtaggaagcaagcgtcttcaagacctgtgtgtatgaatla | 2770 |
| Q | 2761 | CTGATCAAGCAAGGAAGCGGGAACCACTGTGATCTGAGGGAAGGAGACCTCGGTAG | 2820 |
| D | 2771 | ctggaatcaagcaaggaagcggtgaacacacactgtagctcgtgaggaacgagacctcgtgag | 2830 |
| Q | 2821 | TGACAGCATGAAGACAATCTCTATATGCAATATATGCAACATCTACGCCCTCTCAT | 2880 |
| D | 2831 | tgaacagcatgaagacaatctcctaataatgaaatactatgcacgacatctacgcccctcat | 2890 |
| Q | 2881 | GGCAGTATGCTATCTCTGAAAGCCATTTGAGAGAGTTGCTTTGCAAGGGAGCGCTCG | 2940 |
| D | 2891 | ggcagatcatgctatctcttgaaagccattcgagagagttgctcttcgccaagsgcaagctcg | 2950 |
| Q | 2941 | AGCTCTCCCTCCGCGCTGCATACACAGGTTTTTCCGAAGGATCTTCCAGACCCCTATGAATT | 3000 |
| D | 2951 | agctctccctccgctgcatgtagcagagcttcttcgaagagatcccttcgaagccctatgaagt | 3010 |
| Q | 3001 | TTTTGACAGCAACCCACACAGGAGATTCTCAACAGGTTTTTCAAAAGCATGATGAAGT | 3060 |
| D | 3011 | ttttgacagcaaccccaacagsgagtcttccaacaggttcttccaagaacatgatagaagt | 3070 |
| Q | 3061 | TGACGTGCGGCTGCGCTTCCAGGCGGAGATGTTTATCCGAAGCGTTATCCGTGGTTCTT | 3120 |
| D | 3071 | tgaagtgcggtcgcttccaagcgagatgttcaacccaagcgttataccctcggtgtcttc | 3130 |

OY 3121 CTGTCGGGAATGATCGAGAGAGTCTCCGCGTTCCTGTCGAGTGGGGCCCTTGT 3180
 |||||||
 Db 3111 ctgctggtggaatgacgcagagagctccctcgtctcctctgctgagctggcccccttctg 3190
 |||||||
 OY 3181 CATCCTCTTTTCAGTCCGACATTTGTCGAGGGTCCGATTCCGGAGCTGAACGCTCT 3240
 |||||||
 Db 3191 catctctcttcagctcgcacatctgtctcagggctcctgattcggagctgaaagctct 3250
 |||||||
 OY 3241 GGACAAATATCAGCAGTACCTTTCTCTCCCAATCAGCTCAGCATACAGGCGCTTGC 3300
 |||||||
 Db 3251 ggaacatacagcagctacccctctcctccacatacagctcagcatacagggctctgc 3310
 |||||||
 OY 3301 CACCATCAGCGCTCATATAAGGAGAGAGTTCGACANATACAGGAGCATCTCTGA 3360
 |||||||
 Db 3311 cacctcaccgcctcacaataaaggcagaggttctgacagatacagagctgctgga 3370
 |||||||
 OY 3361 TGACAAACCAAGCTCCTTTTCTTTTGTACGTCGTCGATGCGGTGGCTGGCTGGCT 3420
 |||||||
 Db 3371 tgaacaacaagctcctcttcttcttcttcttcttcttcttcttcttcttcttcttct 3430
 |||||||
 OY 3421 GGACCTCATCAGCATGCGCTTCATCAGCAGCAGGCGCTGATGATGCTTTATGCACG 3480
 |||||||
 Db 3431 ggaacctcatcagcatcgccctcatcacacacacagggctgattgcttcttcttcttct 3490
 |||||||
 OY 3481 GCAGATTCGCCAGCGCTATGCGGGCTGCGCCATCTTATGCTGTCCAGTTAACGGGCT 3540
 |||||||
 Db 3491 gcaagattccccagcctctgctggtctgcacatctcttcttcttcttcttcttcttct 3550
 |||||||
 OY 3541 GTTCAGTTTACGTCAGACAGTCGATGAGCAGACAGAGCTCATTCACCTCGTGGAGAG 3600
 |||||||
 Db 3551 gtccaagtttaacgctcagacacggtgactctgagacagaaagctcgaattcactcgtg 3610
 |||||||
 OY 3601 GATCATCATCATATTAAGACTGTGCTTGGAGACACCTGCGCAGAAATTAAAGACAGCG 3660
 |||||||
 Db 3611 gatcaatcactacataaagactctgcttctggaagacactgacagaaatlaagaacaagc 3670
 |||||||
 OY 3661 TCCCTCCCTGACCTGCGCCCGAGAGGAGAGTGAACCTTTGAGAACGCAATAGATGA 3720
 |||||||
 Db 3671 tccctccctctgacgtgccccagagagaggtgacacttctgaaacgcgaagataagga 3730
 |||||||
 OY 3721 CCGAGAAAACCTCCCTCTGTCCTTAAAGAAAGTATCTTACAGATCAAACTTAAGAGA 3780
 |||||||
 Db 3731 ccgagaaaaacccctctctgtctctaaagaagatctccttcaagatcaaaacctaaagaa 3790
 |||||||
 OY 3781 GATTGGCATTTGTGGGGCGAGATCAGGAGATCTCGCTGGGGAGTGGCCCTCTTCG 3840
 |||||||
 Db 3791 gattggcatctggtgggagagacagatcaggaagctcctgctgggagatgycctcttcg 3850
 |||||||
 OY 3841 TCTGTGAGATTATCTGAGAGTGTGATCAAGATTTGATGAGAGTGAATAGATATTGG 3900
 |||||||
 Db 3851 tctgtgtgagttatctgagagctgtcacaagaatctgagtgagagaaatcagtaattgg 3910
 |||||||
 OY 3901 CCTTCCCACTCCGAAACAACCTCTATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3960
 |||||||
 Db 3911 ccttcccgacctccgaagaacactctctctctctctctctctctctctctctctctct 3970
 |||||||
 OY 3961 CACTGTCAATCAAAATTTGGACCCCTTCAACCACTGACACTGAAGACAGATTTGGAGTC 4020
 |||||||
 Db 3971 cactgtcagatcaaaatttggacccctcaaccagtaactgaaagcagatttggatgc 4030
 |||||||
 OY 4021 CCTGAGAGGACACATGAAGAAATGATGCTCAGCTACTCTGAAACTTGAATTTCTA 4080
 |||||||
 Db 4031 cctgagagggagacacacagaaagaaatgattgtctcagctcctcgaacttgaactctga 4090
 |||||||
 OY 4081 AGTGATGAGAAATGGAGTAACCTTCTCAGTGGGAGAACGCGCTTTTGCTAGCTAG 4140
 |||||||
 Db 4091 agtgatgagaaatggagtaacttctcagtggtggaagcagcctctgtgatatgctag 4150
 |||||||
 OY 4141 AGCCTGCTCCGCACTGTAAATTTGATTTTATGATGAAGCACAAGCTGCAATGACAC 4200
 |||||||
 Db 4151 agcctgtctccgcactgtaaagattctgatttagatgaagcacaagctgcacatgacac 4210
 |||||||

OY 4201 AGAGACAGACTTATTGATTCAGAGACCATCCAGAACGATTTGACAGTACATGCT 4260
 |||||||
 Db 4211 agagacagacttattgatttcaagaagacacatccagaaagatttgcagactgtacacatgt 4270
 |||||||
 OY 4261 GACATTTGCCATTCGCTTCGACACAGGTTCTAGGCTCCGATAGATTTAGTCTGGCCCA 4320
 |||||||
 Db 4271 gacatttgccatcgcctgcgcacacagcttctagctcccgtaagattatgtgtgctgccca 4330
 |||||||
 OY 4321 GGGACAGGAGGAGTGTGAGACACCCCATCGCTCTTCTGTCCAAACGACAGTCCGAT 4380
 |||||||
 Db 4331 gggacagaggtgtgagtttgagacaccccatcgctcctctgtccaaacgaaagttcccgatt 4390
 |||||||
 OY 4381 CTATGCAATGTTGGTGGCTGCGAGAACAAAGTCCGTGCAAGGCGTCACTCCCTGT 4440
 |||||||
 Db 4391 ctatgcaatgttggctgtcgcagagaaacaaggtcgtctcaagggctgactcctccgt 4450
 |||||||
 OY 4441 TGACGAAGTCTCTTTTCTTTTATAGAGCATTCGCAATTCCTGCGGGCGGGCCCTTCATC 4500
 |||||||
 Db 4451 tgaagaagctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct 4509
 |||||||
 OY 4501 GCGTCTCTACCGAAACCTTGCCTTTCGATTTTATCTTTCGACAGCATTCGGAT 4560
 |||||||
 Db 4510 gctctcctcaccgaacacttgccttctcgaatttcttcttcttcttcttcttcttcttct 4569
 |||||||
 OY 4561 TGGCTGTGTGTTTCACTTTTAGGAGAGTCATATTGATTGATTATTTTCATAT 4620
 |||||||
 Db 4570 tggctgtgtgttcaactttagggagagtcataattgtatcttcttcttcttcttcttct 4629
 |||||||
 OY 4621 TCATGTAACAAATTTAGTTTCTTTTCTTAAATGCACTTAAAGGTTCAAGGAACCT 4680
 |||||||
 Db 4630 tcatgtaaacaaattttagtttcttcttcttcttcttcttcttcttcttcttcttcttct 4689
 |||||||
 OY 4681 TATTTAATTTATCAGAGGCTTATATATACGTTTATACGTTATATCTATATAT 4740
 |||||||
 Db 4690 tatttaatttctacagagccataatgaacttatacgttgcgtatatactataata 4749
 |||||||
 OY 4741 AATTCGTATAGCCTATATTTTACAGTAAGTAAGTAAAGCTGTTATTTATTTAAAT 4800
 |||||||
 Db 4750 aatctgtacataagccataatcagtgaaatgaagctgttcttcttcttcttcttct 4809
 |||||||
 OY 4801 AAGCACTGTGCTAAAAAA 4819
 |||||||
 Db 4810 aagcactgtgtcaataaca 4828
 |||||||

RESULT 4
 AA294745
 ID AA294745 standard; cDNA; 5838 BP.
 XX
 AC AA294745;
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Human ATP binding cassette ABC5 (MRP5) cDNA.
 XX
 KW ABC5; ATP binding cassette; human; cholesterol; lipid disorder;
 KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
 KW lupus erythematosus; diagnosis; gene therapy; MRP5;
 KW multidrug resistance associated protein; chromosome 3q25-27;
 KW Dublin Johnson syndrome; hyperbilirubinemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200018912-A2.
 XX
 PD 06-APR-2000.
 XX
 PF 21-SEP-1999; 99WO-EP06991.
 XX
 PR 25-SEP-1998; 98US-0101706.
 XX
 PA (FARB) BAYER AG.

PI Schmitz G, Klucken J;
XX WPI; 2000-293151/25.
XX Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders -
XX
XX Claim 9; Page 123-125; 154pp; English.
XX
XX The present sequence is that of human ATP binding cassette
CC subfamily C protein ABC5 cDNA. The cDNA was identified using a
CC differential display method in which monocytes from peripheral
CC blood were subjected to macrophage differentiation and cholesterol
CC loading with acetylated low density lipoproteins and subsequent
CC degrading with high density lipoprotein (HDL3) to identify
CC cholesterol sensitive genes. The gene maps to chromosome 3q25-27
CC and is also termed MRP5 (multidrug resistance associated protein).
CC This is the gene locus for Dublin-Johnson syndrome, a disorder
CC associated with mild chronic conjugated hyperbilirubinemia.
CC The invention provides cholesterol-sensitive ABC genes (see
CC AA294734-63). These genes, and polypeptides encoded by them,
CC can be used for diagnostic and therapeutic applications, and for
CC biochemical or cell-based assays to screen for pharmacologically
CC active modulator compounds useful for the treatment of lipid
CC disorders, atherosclerosis or other inflammatory diseases such as
CC psoriasis and lupus erythematosus.
XX
XX Sequence 5838 BP; 1422 A; 1432 C; 1509 G; 1475 T; 0 other;
SQ

Query Match 98.98; Score 4792.6; DB 21; Length 5838;
Best Local Similarity 99.88; Pred. No. 0;
Matches 4809; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1 GGCCTCATGCTCGGGAGGCTGGTGTAGGGGCGGCGGTTGCTCGAGCAGGGGGCGAG 60
DB 11 ggcctcatgctcgggagggctgggtgtagggcgcggttgctcgtgagcggggcgag 70
QY 61 GAATTCGATGTGAATACAGTCTGTGAGCCTGTGAACTCCACACAGAGAGATGAA 120
DB 71 gaattcgaatgtgaatacagctctgtgagcctgtgaaactccacacagagagatgaa 130
QY 121 GGATATGACATAGGAAAGAGTATATCATCCCAAGTCTGGTATAGAGTGTAGGGA 180
DB 131 ggaatgacataggaagagatatacatcccaagtctggatagagtgtaggga 190
QY 181 GAGAACGACAGCTCTGGAGCGCAGACAGACCGGAGATGCCAAGTCCAGAGAACTCG 240
DB 191 gagaacgacagctctggagcgcagacagacccggagatgcccaagtccaggaactcg 250
QY 241 ACCGTTGGAATGCCAAGATGCTTGAAGACAGCCGAGCCGAGGGCCTCTCTTGA 300
DB 251 accgttggaatgccaaagatgcttgaagacagcccgagcccgagggcctctcttga 310
QY 301 TGCCTCATGCAATCTCAGCTCAGATCTCGAGAGAGAGCATCCCAAGGAAATACCA 360
DB 311 tgcctcatgcaatctcagctcagatctcgagagagagcatcccaaggaaatacca 370
QY 361 TCATAGGCTTAGTGTGAGAGCCCATCCGACACTTGTGCAACACAGCCCAAGTGA 420
DB 371 tcataggcttagtgtgtagagcccatccgacacttgtgcaaacacagcccaagtga 430
QY 421 CAATGCTGGGCTTTTCTCGTATGATGATTTTCTGTGGCTTCTCTGCGCCGTGAGC 480
DB 431 caatgctgggcttttctcgtatgatgatttctgtggcttctctgcccgtgagc 490
QY 481 CCACAGAGAGGGAGGCTCTCAATGAGAGAGTGTGCTTCTGCAACAGCAGATCTTC 540
DB 491 ccacagagagggagggctctcaatgagagagtggttctgcaaacagcagatcttc 550
QY 541 TGACGTAACTGCAAGAGACTAGAGACATGTGCAAGAGAGAGCTGAATGAATGGGCC 600
DB 551 tgacgtaaactgcaagagactagagacatgtgcaagagagagctgaatgaatgggcc 610

QY 601 AGAGCTGCTTCCCTGCGAAGGTTGTGTGATCTTGTGCGACACAGGCTCATCTGTC 660
DB 611 agagctgcttccctgcgagggttgtgtgatcttgtgctgcgcacacaggtccatctgctc 670
QY 661 CATGTGTGCTGATGATACAGCAGCTGCTGCTTCAAGTGGACAGCTTCAAGTGA 720
DB 671 catgtgtgctgatgatacagcagctgctgcttcaagtggacccttcaagtga 730
QY 721 ACACCTCTGGAGTATACAGCAGAGAGCTAATCCGACAGTACAGCTGTGTAGT 780
DB 731 acacctctggagtatacagcagagagcttaacctgagctgagcttctgttagt 790
QY 781 GCTGGGCTCTCTGACGAGAAATGTCGGTGTGTGCTGCTGCACTGACTGGGCAAT 840
DB 791 gctgggctctctgacgagaaatgtcgggtgtgtgctgctgcaactgactgggcaat 850
QY 841 GAATTCAGCAACCGGTGCTGCGGGGGCCATCTTAACATGAGATTTAAGAAAT 900
DB 851 gaattcagcaaccgggtgctgcgggggccatcttaacatgagatTTAAGAAAT 910
QY 901 CCTTAAGTAAAGAACATTAAGAGAAATCCCTGGTGTGAGCTCATCAATTTGCTCAA 960
DB 911 ccttaagttaaagaaacatTAAGAGAAATccctgggtgtgagctcatcaatTTGCTCAA 970
QY 961 CGATGGCAGAGAAATGTTGAGCAGCAGCCGTTGGCAGCTGCTGCTGAGAGACCCGT 1020
DB 971 cgatggcagagaaatgTTGAGCAGCAGccgTTGGCAGCTGCTGCTGAGAGACCCGT 1030
QY 1021 TGTGGCATCTTAGCATGATTTATATGTAATTTGTGGGACCAACAGCTCTCTGAG 1080
DB 1031 tgtggcatcttagcatgatttataTATGTAATTTGTGGGACCAACAGCTCTCTGAG 1090
QY 1081 ATCAGCTGTTTATACCTTTTACCCAGCATGATGTTTCATCAGCGCTCACAGCATA 1140
DB 1091 atcagctgttttatacccttttaccagcatgattgtttcatcagcgctcacagcata 1150
QY 1141 TTTGAGGAAATGCGTGGCCGCCACGAGTGAACGTGTCCAGAAAGATGAATGAA 1200
DB 1151 tttagaggaaatgCGTGGCCGCCACGAGTGAACGTGTCCAGAAAGATGAATGAA 1210
QY 1201 TACTTACCTTAATTTATCAAAATGTATGCTGGGTCAACACTTTTCTAGAGTGTCA 1260
DB 1211 tacttaccttaatttatacaaaatgtatgctgggtcaaacactTTTCTAGAGTGTCA 1270
QY 1261 GAAATCCGCGAGGAGGAGCGTGGATTTGAAAGAAACCCGGTACTCCAGAGCATCAC 1320
DB 1271 gaaatccgCGAGGAGGAGCGTGGATTTGAAAGAAACCCGGTACTCCAGAGCATCAC 1330
QY 1321 TGTGGGTGTGCTCCCATTTGTGTGTGATTTGCCAGCGTGTGACCTTCTGTTCATAT 1380
DB 1331 tgtgggtgtgctcccatTTGTGTGTGATTTGCCAGCGTGTGACCTTCTGTTCATAT 1390
QY 1381 GACCCGTGGCTTCATGATGACAGCAGCAGGCTTTTCAACATGTTGACAGCTTCAATTC 1440
DB 1391 gacctgtggcttcATGATGACAGCAGCAGGCTTTTCAACATGTTGACAGCTTCAATTC 1450
QY 1441 CATGACTTGTGCTTGAAGTACACCGTTTCAAGTAAAGCCCTCGAAGAGCTCAGT 1500
DB 1451 catgacttGTGCTTGAAGTACACCGTTTCAAGTAAAGCCCTCGAAGAGCTCAGT 1510
QY 1501 GGCCTGTGACAGATTTAAGATTTGTTTCAATGAGAGAGTTCACATGATTAAGAACAA 1560
DB 1511 ggcctgtgacagattTAAGATTTGTTTCAATGAGAGAGTTCACATGATTAAGAACAA 1570
QY 1561 ACCAGCCAGTCTCATCATCAAGATAGAGATGAAGAAATCCACTTGGAGTGGACTCTC 1620
DB 1571 accagccagtctcatcatCAAGATAGAGATGAAGAAATCCACTTGGAGTGGACTCTC 1630
QY 1621 CCATCTCAGTATCCAGAACTGCCCCAAGCTGACCCCAATTAAGTAAAGAGCAGAGGCG 1680
DB 1631 ccactctcagtatccagaaactgCCCCAAGCTGACCCCAATTAAGTAAAGAGCAGAGGCG 1690

```

1681 TTCCAGGGGCAAGAAAGAGAGTGAGGCGAGCTGCAGCGCACTGAGCATCAGCGGTGCT 1740
1691 ttccaggggcaagaagaagaggtgtaggcagctgcagcgcactgagcatcagggcgtgct 1750
1711 GGCAGAGCAAGAAAGGCCACCTCTCTCTGAGACGTGACGAGCGGCCAGTCCGGAAGAGA 1800
1751 ggcagagcaagaagggccacctctctctctgtagcagtagcagcgcccaagccgaagaaga 1810
1801 AGAAGGCAAGGACATCCACACCTGGGCGACCTGGCGCTTACAGAGACACACACATCGA 1860
1811 agaaaggcaagcacatccaccctgggcccactgcgtcttaacagaggaaactgcacagatcga 1870
1861 TCTGAGATGCCAAGAGGGTAAACTGTTGSAATCGCGCACTGCGAAGTGGAAAAAC 1920
1871 tctgagatcccaaggggttaaacttgtagaacctgcgcgagtgtaggaaagtcggaanaaa 1930
1921 CTCTCTCATTTTCAGCCATTTTAAAGCCAGATGACGCTTCTAGAGGCGACATTCGATCAG 1980
1931 ctctctcatcttaagccattttagccagatgagcgtctctaaggggcagcatctgcaatcag 1990
1981 TGGAGCCTTCGCTTATGTGTGGCCCGACAGCGCTCGATCCATGCTACTCTGAGAGACAA 2040
1991 tggaaaccttcgcttatgtggtggccagcgctgtagtccctcaatgctactctgagagacaa 2050
2041 CATCCTGTTTGGAGAGATATGATGAAGAAAGATACAACTGTCGTGAACAGCTGCTG 2100
2051 catcctgtttggagagataatgtagaagaaagatacaaaactctgtcgtgaacagctgctg 2110
2101 CCTGAGGCGCTGACCTGGCCATTCCTCCAGCAGCGACCTGACGAGGATTTGAGAGCAGG 2160
2111 cctgagcgctgacctggccatcttccagcagcgacctgagcgagatctggagcgag 2170
2161 AGCCAACTGAGCGGTGGGCGACGCCAGAGATCAGCCTTGCCCGGCTTGTATATGTA 2220
2171 agccaaactgagcggtggcgcgccagagatcagcgtctgcgcgggctctgatatagta 2230
2221 CAGGACATCTTCATCTCGAGAGACGCCCTCAGTGCCTTAGTGCATGTCGGAACCA 2280
2231 cagggacatctacatccctggagcgacccctcagctgagcttagtgcccatgtggtgaacca 2290
2281 CATCTTCATTAAGTGTATCCGGAACATCTCAAGTCCAAAGACAGTTCGTTGTTAACCA 2340
2291 catctcaatagtgctalcgcggaacatctcaagtcgaagacagctctgtgttaacca 2350
2341 CCAGTTAAGATACCTCGGTGAGCTGTGATGAGTGAATCTTATGAAAGAGCGCTGATTAAC 2400
2351 ccagttacaagtaacctgtgtagctgtagaagtgatcttcatcgaagagggcgtgattaac 2410
2401 GGAAGAGGCAACCATGAGAGAACTGATGAATTAATGGTGAATGCTACCATTTTAA 2460
2411 ggaagaggcacccatgaggaactgtagaacttaaaatggtgactatgctaacatctttaa 2470
2461 TAACTGTTGCTGGGAGAGACACCGCCAGTTGAGATCAATTCAAAAAAGGAACCACTGG 2520
2471 taacctgttgcgggagagagacccgcgctgtagatcaaatcaaaaaaaggaaccaagtg 2530
2521 TTCAACAGAAAGATCAACAGACAGGCTCTTAAACAGGATCAATPAAGACGAAAAAC 2580
2531 ttcaacaagaagaagtcacaagaacgaaggttctaacaacagatcagtaaaagaagaaagc 2590
2581 AGTAAGGCAGAGGAAGGGGAGCTGTGACGTGGAAGAAAGGAGGAGGCTTCAGTGCC 2640
2591 agtaaaagcagaaggaagggcagctgtgcagctggaagaaaggaagggcaggttcagtgcc 2650
2641 CTGCTCATATATGATGTGATACATTCAGGCTGCTGGGGCCCTTGCGATCTCTGTTAT 2700
2651 ctgctcatatattggtgtctataatccagcgctgtagggcccttggaatccctgtagtat 2710
2701 TATGGCCCTTTTCAATGCTGAATGATGAGCAGACCGCCTTCAGCACTGGTGGTTAGTTA 2760
2711 tatggcccttttcaatgctgaatgtagagcagccgcttcagcacctgtagttagttagtt 2770
2761 CTGATCAAGCAAGGAAGGGAACCAACCTGTGACTGAGGGAACGAGACTCGGTGAG 2820
2771 ctgatacgaaggaagcggggaacaacaactgtagctcgagggaacgcgactcgtgag 2830
2821 TGACAGCATGAAGACATCTCTATATGACATGACTATGTCACACATCTACGCCCTTCAT 2880
2831 tgacagcatgaagacaatccctacataatgagtaactatgcagcatctaacgacctctcat 2890
2881 GGCAGTCATGTCGATCCGTAAGGCCATTCGAGAGTGTGCTTGTCAAGGCGACGCTGGC 2940
2891 ggcagtcagtcagatcccgaaagccatcgaagagtgctgtctgttcaaggaagcagctgcg 2950
2941 AGCTTCCTCCCGCTGACATGACAGACGCTTTCCGAAGGATCTTCCAAAGCCCTATGAAT 3000
2951 agcttcctcccgctgacatgacgagcgtttccgaagagatcccttcgaagccctatgaagt 3010
3001 TTTTACACGAGACCCCCACAGAGGAGATTCACAGAGTTTTCAAAAGACATGATGAAGT 3060
3011 ttttacaagacccccaagggagagattccaacaggttttccaagaacatgtagaagt 3070
3061 TGACGTMGGGCTGCCGTTCCAGCGGAGATGTTCAATCCAGAGTTATCTGTGTTCT 3120
3071 tgaagtggtgcgcgtctccagggccgaagatgttcaaccaagacgtatccgtgtctct 3130
3121 CTGTGTGGGAATGATCGCAGAGATCTCCGATGCTTCTTGTGGCAGTGGGCGCCCTTGT 3180
3131 ctgtgtgggaatgatacgagagatctccggtgcttccgtgtctctgtggaagtgggccccctgt 3190
3181 CATCTCTTTTCAGTCTGACACATCTTCACAGGCTCTGATTCGGAAGCTGAAGCGTCT 3240
3191 catctcttttcagtcctgcacatgcttccaggttccaggttccgtatctcggaagctgaagcgtct 3250
3241 GGACATATACAGCAGTACCTTCTCTCCACATCAGCTCAGCATTCAGAGGCGCTTGC 3300
3251 ggaacatacaacgacgacacacttccctccacacacagctccacacacacacagggccttgc 3310
3301 CACCATCCACGCGCTCAATTAAGAGGCGAGAGTTCGACAGATATCCAGAGACCTCTGA 3360
3311 caccatccacgctcaacaataaagggcagaggttctgcacagaataccagagctgtgga 3370
3361 TGACAAACCAAGCTCTTTTGTGTTTACGTGTGCGATGCGGTGCGTGTGCGGCT 3420
3371 tgaacaacaagctctcttttttctgtttagctgtgtagtggtgtgctgtgcgtgcgt 3430
3421 GGACCTCATACAGATCGCCCTCATACACACAGGGGCTGAGTACGTTATATGACAGG 3480
3431 ggaactcaacagcatcgccctcatcaaccaaccaagggcgtgtagatcgtctttagcagcg 3490
3481 GCAGATTCGCCAGCGCTATGCGGATCGCATCTCTTATGCTGTGTCAGATTAAAGGGGCT 3540
3491 gcaagatcccccaagcctatggtgtgtcgtgcacatccctatgctgtccagttaaaggggct 3550
3541 GTTCCAGTTTACGTCAGACTGAGCATCTGAGACAGAAAGCTCGATTCACCTCGGTGAGAG 3600
3551 gtccagtttaagtcagagctgtgcatcgtgagacagaagctcgaatccaacctgtgtaggg 3610
3601 GATCAATCACTACATTAAGACTGTCTCTTGAAGCAGCTGCGCAAAATTAACAAAGGC 3660
3611 gatcaatcaactacatlaagaactgtctctggaagacacctgcagaaatlaaagaacagggc 3670
3661 TCCCTCCCTGACTGCGGCCACAGAGGAGAGTGAACCTTTGAACGCGCAGATGAGGTA 3720
3671 tccctccctgactggtgcccagaaggaaggtgacaccttggaacgcagagatgtaggta 3730
3721 CCGAGAAAACCTCCCTCTGCTGCTTAAAGAAATATCTTCACAGATCAAACTPAAGAGAA 3780
3731 ccgagaaaaacctccctctgtcttaagaagatattccttcaagatcaaaacctaaagagaa 3790
3781 GATTGGCATTTGCGGGCGGAGACAGGATCAAGGAAAGTCTGCGTGGGGATGCGCCTCTTCCG 3840
3791 gatggcatttggtgggcggaacagatcaaggaaggtctgctggggatgagcccttccg 3850
3841 TCTGTGAGTATCTGAGAGGCTGCATCAAGATTGATGGAGGAGAAATCAGATATTGG 3900
```

```
Db 3851 tctggtgagattatctgaggctgcatcaagaattgatgagtgagaatcaagtatttg 3910
QY 3901 CTTTGGCCGACCTCCGAAGCAAACTCTCTATCATCTCTCAAGAGCGGTGCTTCACTGG 3960
Db 3911 ccttgccgaactccgaagcaaacctctcatctatctcccaagagccggtgctgctcagtg 3970
QY 3961 CACTGTGATCAATTTTGGACCCCTTCAACAGTACAGTACAGACCAAGATTGGATGC 4020
Db 3971 cactgtaaatcaatttggacccctcaaccagtaacatgaaagacagatttggatgc 4030
QY 4021 CCTGTGAGAGACACACATGAAAGAAATGATTCAGCTACCTGAACTGATGATGCA 4080
Db 4031 cctggaagagacacacatgaaagaaatglatgtcagctacccctgaaacctgaaatcga 4090
QY 4081 AGTGATGAGAAATGGGATACTTCTCAGTGGGGAGACGGCAGCTCTGTGTCATAGCTAG 4140
Db 4091 agtgaaggaaatgaggaaactctcagtgagggaacgagctcttgctcagaagctag 4150
QY 4141 AGCCCTGCTCCGCCACTGTGAAGATTTCGATTTCGATGAGAGCCACAGCTGCCATGACAC 4200
Db 4151 agccctgtccgcacatgtaagattctgatttagatgaagccacagctgcacatggaac 4210
QY 4201 AGAGACAGACTTATTTGATTCAGAGACATCCGAGAAAGCATTTGACAGCTTACATGCT 4260
Db 4211 agagacagacttattgtatccaagagacatccgaaagacattgcagacatgacaagctc 4270
QY 4261 GACCATTTGCCATGCGCTGCACACAGGTTCTAGGCTCCGATAGATTATGATGTCGCCCA 4320
Db 4271 gacattgcccactgcctgcacacagctctcagctcgaatgattatgctgcgagccca 4330
QY 4321 GGGCAGAGTGGTGGAGTTTGACCCCATGCTCTTCTGTCACACACATTTCCCATTT 4380
Db 4331 gggacaggttgtagtatttgacaccccatcgctcctctgccaacagatcccgat 4390
QY 4381 CTATGCCATGTTTCTCTGCTGCAGAGAAACAAGTGCCTGTCAAGGGCTGACCTCCCTGT 4440
Db 4391 ctatgcatgttctgctgctgtagagaaagctgcgtcgaagggctgcacccctccgt 4450
QY 4441 TGACGAAGTCTCTTTCTTTTGAAGACATTTGCCATTCCTGCTGGGGGGCCCTTCATC 4500
Db 4451 tgaagaagctctcttctctttagagcatgcatccctgcctggggggggccctcctc 4509
QY 4501 GCGCTGCTCTGCGAAACCTTGCTTCTGATTTTATCTTTCGACACAGCTCCGGAT 4560
Db 4510 gcgctccctcaacgaacactgctctctcgaatttactcttcgaagcagctccgat 4569
QY 4561 TGCTTGTGTTTCACTTTAGAGAGATCATATTTTGATTTATTTATTTCCATAT 4620
Db 4570 tggcttggtgcttcaactttagggagatcatatttgatattatattatccatat 4629
QY 4621 TCATGTAAACAAATTTAGTTTGTCTTATTTGACCTTAAAGGTTCAGGACCCGT 4680
Db 4630 tcatgtaaacaaatttagtttcttcttactatgcatcctcaaaagttcagggaacgct 4689
QY 4681 TATATATATTTATCAGAGGCTTATTAATGAAGCTTATAGCTAGCATTTATATAT 4740
Db 4690 tatataatgtlacaagagcctaataagccttacaagctgtagcatacataat 4749
QY 4741 AATTTCTATACATACCTATATTTTACAGTGAAGAAATGAAGCTTTATTTTATTTAAAT 4800
Db 4750 aattctgtacatagcctatatattacagtgaaactgtaagcgttattattatataaat 4809
QY 4801 AAGCACTGTGTAATAAAA 4819
Db 4810 aagcactgtgtaataaca 4828
```

```
RESULT 5
AAC85287 standard; cDNA; 4781 BP.
xx
AC AAC85287;
xx
```

```
DT 29-MAR-2001 (first entry)
DE Multidrug-resistance associated polypeptide-beta coding sequence.
XX
KW Multidrug resistance-associated polypeptide; MRP-beta; transporter;
XX P-glycoprotein; chemotherapy; cancer; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2..4363
FT /tag= a
FT /product= "MRP-beta"
XX
PN US6162616-A.
PN
XX
PD 19-DEC-2000.
XX
PF 16-APR-1997; 97US-0843459.
XX
PR 16-APR-1997; 97US-0843459.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PI
XX Shyjan A;
XX
DR WPI; 2001-111728/12.
DR P-PSDB; AAB47021.
XX
PT Novel nucleic acid encoding multidrug resistance-associated protein
PT MRP-beta, useful for detecting and treating drug-resistant cancers -
XX
PS Claim 1; Fig 1; 41pp; English.
XX
CC This sequence encodes a multidrug resistance-associated polypeptide
CC (MRP-beta). MRP-beta is a transporter that can remove
CC chemotherapeutic agents from cells and/or sequester such agents,
CC and it probably accounts for multidrug resistant phenotypes that do
CC not respond to treatments with reversal agents directed against
CC P-glycoprotein or the known MRP. By inactivating MRP-beta, at nucleic
CC acid or protein levels, cells should be rendered sensitive to
CC chemotherapy agents. The MRP-beta coding sequence, or fragments
CC of it, can be used to detect expression of the MRP-beta gene and
CC thus multidrug resistant cells. Fragments of MRP-beta DNA are useful
CC as primers and as antisense (therapeutic) agents for inhibiting
CC expression of the MRP-beta gene. They can also be used for detecting
CC mutations and to treat diseases associated with abnormal MRP-beta
CC gene expression, specifically cancer, particularly for potentiating
CC chemotherapy of fluid or inoperable tumors, or in cases of regrowth
CC after an initial course of chemotherapy.
XX
SQ Sequence 4781 BP; 1217 A; 1161 C; 1232 G; 1165 T; 6 other;
XX
Query Match 98.5%; Score 4775.4; DB 22; Length 4781;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4773; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 67 TGAGTGAACATAACAGTCTGTGAGCCCTGGAGACCTCACTCAGAGAAGATGAAGATAT 126
Db 1 tgaatgaactaactacagctgtgagccctggaactctcaactcagaagaatgaaagatat 60
QY 127 CGACATAGGAAAGAGATATCATCCCAAGTCTGGGTATTAAGAGTGAAGGAGAGAAC 186
Db 61 cgacataggaagaagatatacatcaccagctccctgggtatagaagtgtgaggagagaac 120
QY 187 CAGCACTTGTGGAGCGACAGAGACCGTGAAGATTCCAAGTTCAGAGAACTGCACCGTT 246
Db 121 cagcactcttgagcagacagacgctgaaagattccaagttcaggaagactgcacgct 180
QY 247 GGAATGCCAAGATGCTTTGAAACAGCAGCCCGAGCCGAGGGCTCTCTCTTGAATGCTC 306
Db 181 ggaatgccaaagatgaccttgaaacagacagccgagccgagggcctctctcttgatgcttc 240
```

| | | | |
|----|------|---|------|
| QY | 307 | CATCATCTTCAGAGTCAGAAATCCTGGATGAGGAGCATCCCAAGGGAAAGTACCATCATG | 366 |
| DB | 241 | catgcatcttcagagctcgaatccttgatigagagcatcccaagggaagatacatcatg | 300 |
| QY | 367 | CTTAGTGTCTTGAAAGCCATCCGGACTACTTGCAAAACACGACACCAAGTGGACATGC | 426 |
| DB | 301 | cttagtgctctgaagcccatccgagctactctccaaacacgacccagctggaataatgc | 360 |
| QY | 427 | TGGGCTTTTTCGATATGACTTTTTCGAGGGCTTCTCTCGGGCCGCTGGGGCCCAAA | 486 |
| DB | 361 | tgggctttttcggatagcttttttcgagggcttctctcgcgcgctggtggcccaaa | 420 |
| QY | 487 | GAAAGGGGAGAGCTCTCATGTGAAGACGCTGTGTCTCTGTCTCAAGACGAGTCTTCGAGCT | 546 |
| DB | 421 | gaagggggagagctccaatggaagacgtgtgtctctctgtccaagcagcggtctctgaagct | 480 |
| QY | 547 | GAACTGCAGAAAGACTAGAGAGACTGTGGCAAGAGAGCTGAATGAATGGGCCAGAGCC | 606 |
| DB | 481 | gaactgcagaagagctagagagactgtgcaagaagagctgaatgaattggccagagcgc | 540 |
| QY | 607 | TGCTTCCCTCGGAAAGGGTTGTGTGATATCTTTCGCGACACAGAGCTCATCTGTCCATCTG | 666 |
| DB | 541 | tgcttccctcgcgaagggtttgtgtgactcttcgcgcgacacaggtccatccctgtccatcgt | 600 |
| QY | 667 | GTGCGCTGATGATCAGCGACGCTGTGCTGTCTAGTGGAGCAGCCTTCATGTGTAAACACT | 726 |
| DB | 601 | gtgcctgatgatcaagcagagctgtgtcttcagtggaacagcttcatatgtgaaacacct | 660 |
| QY | 727 | CTTGGAGTATACCCAGGCAACAGAGTCTAACTGCAGTACAGCTTGTGTGTAGTGTGGG | 786 |
| DB | 661 | cttggagtatacccaagcaacagagctaaacctgcagctcagctgtgtagtgctg | 720 |
| QY | 787 | CCCTCCTCCTGACGGAAATCGTGGGGGCTTGGTGGCTGTGACGACTTGAGGCAATTGAATA | 846 |
| DB | 721 | ccctccctcgcagcggaaatcgtgtgcctctgtgtcgtctgcactgcagcttgggcaattgaata | 780 |
| QY | 847 | CCGAAACGGGTGCTCCGCTTGGGGGGCCCAATCCCTAAACAAGGCAATTGAAGAGATCTTAA | 906 |
| DB | 781 | cggaaacgggtgtccgcttggggggccatccctaaccatgagatattaagaagatccttaa | 840 |
| QY | 907 | GTTTAAAGACATTAAAGAAATCCTGGGTGAGCTCATCAACATTTGCTCCACAGATGG | 966 |
| DB | 841 | gtttaaagacaattaaagaatacccttgggtgagctcatcaacaattgtccaaagatlg | 900 |
| QY | 967 | GCAAGAGATGTTTGAAGGCACACACGCTGTGGAGCGCTGTGAGAGACCGGTGTGGC | 1026 |
| DB | 901 | gcagagagaattgttgaagcagcagcgttgcagcctgcgtctgtgaaggaaacctgtgtgc | 960 |
| QY | 1027 | CATCTTAGGCGATGATTTATTAATTAATTAATTAATTCGTGGACCAAGAGCTTCGTGGATCAGC | 1086 |
| DB | 961 | catctttagcagtattatataatgtaatattctctggagccaagaagcttcccttggagtcgcg | 1020 |
| QY | 1087 | TGTTTTTATCCTCTTTTATCCACAGCAATGATGTTTGCATCAGGCTCACAGCATATTTTCAG | 1146 |
| DB | 1021 | tgttttatccctctttatcccgcaatgattgttgatcaogcgtccaaagcatatttcag | 1080 |
| QY | 1147 | GAGAAATGCGTGGCCGCCACGAGTGAAGCTGTGCCAAGAAAGATGAATGAATTCTTACTTA | 1206 |
| DB | 1081 | gagaaaatgctgtgcccgcacagatgaacgtgtgccagaagatgaatgaagtcttactta | 1140 |
| QY | 1207 | CATTAAATTTATCAAAATGTATGATCGTGGGCAAGACATTTTCTCGAGAGGTTCAACAATAT | 1266 |
| DB | 1141 | cattaaatttatacaaatgtatgacgtcgggtgcaagaatcttcttcagagtggttcagaatat | 1200 |
| QY | 1267 | CCGAGAGAGAGAGCGTGCATTTATTTGAAAAAGCCGGTACTTCCAGAGACATCACTGTGG | 1326 |
| DB | 1201 | ccgagagagagagtcgtgatatttgaaaaagccgggtactcttcagagatcatcgtgtgg | 1260 |
| QY | 1327 | TGTGGCTCCATTGTGTGTGTATGTGCCAGCGTGTGACCTTCTCTGTTCATATGACCTT | 1386 |
| DB | 1261 | tgtggtcccatcttgtgtgtgtatgtccagcgtgtgagacctctctgttcatataacct | 1320 |

| | | | |
|----|------|--|------|
| QY | 1387 | GGGCTTGATCTGCACGACACACAGGCTTTACACAGTGGTGAACAGCTTTCATTTGCATGAC | 1446 |
| Db | 1321 | gggcttcgatctgcacgacgacacaggctcttcacagtgctgacagcttccaattccattgac | 1380 |
| QY | 1447 | TTTTGGCTTGAACCTACACCGTTTTTCAGTAAAGTCCCTTCAGAAAGGCTCAGTGGCGGT | 1506 |
| Db | 1381 | tttgccttgaagaataacaccgctttccagaaagtcctccctcagaagcccaagtgctgcgt | 1440 |
| QY | 1507 | TGACAGATTAAAGATTGTTTCTAAATGGAAGAGGTTCACATGATAAAGAACCAACACAGC | 1566 |
| Db | 1441 | tgacagattaaaggttgccttctaataatggaagaggttcacatgatataaagaacaacacagc | 1500 |
| QY | 1567 | CAGTCTTCATCAGATAGAGATGAAAAATGCCACTTGGCATGGACCTCTCCCTACTC | 1626 |
| Db | 1501 | cagtcctcaatcaagaatagaatgaaaaatgcacccttgcatgagactcctccactc | 1560 |
| QY | 1627 | CAGTATCCAGAACTCGGCCAAGCTGACCCCCAAAATGAAAAAGAACAGAGGGGCTTCAG | 1686 |
| Db | 1561 | cagtatccagaacatcgcgccaagctgcagcccccaaaatgaaaaagaacaagagggcttcag | 1620 |
| QY | 1687 | GGGCAAGAAAGAGAGGTGAGGCGATGACGCGCATCTGAGCTACGAGGCGTGTGGCAGA | 1746 |
| Db | 1621 | gggcaagaagaagagagtgagtcgagctgcagcgcactgagcatcaagcggtgctgcaga | 1680 |
| QY | 1747 | GCAGAAAGGCCACTCTCTCTGTGACAGTGCAGCGGGCCCACTCCCGAAGAGGAAGAAG | 1806 |
| Db | 1681 | gcagaaagggccactctcctctctgacagtgacgagcgccagctcccgaaaggaagaag | 1740 |
| QY | 1807 | CAGGCACATCCACTGGGCGCACTGGGCTTACAGAGAACATGCAACAGATCGATCTGGA | 1866 |
| Db | 1741 | caaggaacatccacctgycgcaactgcgcttaacagagaaacgcgcacgacatgcactgga | 1800 |
| QY | 1867 | GATCCAAAGAGGATTAACGTGTGGAATCTGCGGACATGTGGGAAATGGAATAAACCTCTCT | 1926 |
| Db | 1801 | gatccaaagaggttaactgcttggaatctgcgcgagtgctgaggaaatggaataaacctcct | 1860 |
| QY | 1927 | CATTTCAGCCATTTTAAAGCCACAGATGACGCTTCTAGAGGCGACGATTCACATCAGTGAAC | 1986 |
| Db | 1861 | catttcagccattttaagccaagatgcagctctctagaaggcagcatgcaatgcagtggaac | 1920 |
| QY | 1987 | CTTGCGTTATGTGGCCGACGAGCGCTGGATCCCATGATGTCGTAGACGACCAACATCTCT | 2046 |
| Db | 1921 | cttcgcttaatgtyggccagagaagccctggaacccccaatgctaacctcgaagaaacaatcc | 1980 |
| QY | 2047 | GTTTGGGGAAGAAATATGATGAAAGAAAGATACACTGTGTGCTGAACAGCTGCTGCCATG | 2106 |
| Db | 1981 | gtttgggaagaataataatgaatgaagaagaatacaactctgctgcgaacagctgctgcctgag | 2040 |
| QY | 2107 | GCGTCGACCTGGCCATTCTTCCACGACGACGCACTGACGAGATTTGAGACGAGAGCCAA | 2166 |
| Db | 2041 | gctgcgacctgycatctctctccagcagcgacctgcagagatctggaagcgagggccaa | 2100 |
| QY | 2167 | CCTGAGCGGTGGGACGAGGCCAGAGAGATCAGACCTTGGCCGGGCTTGATATGACAGAGAG | 2226 |
| Db | 2101 | cctgagcggttgycagcgcagagagatcaagcctctgcgcgggctctgtataatgacaagag | 2160 |
| QY | 2227 | CATCTACATCCCTGGACGACGCCCTCAGTGGCTTAAATGTCCTAATGTCGGAACACATCTT | 2286 |
| Db | 2161 | catctaacctccctgcgaacccccctcagctgccttaagatgcacatgtygggaacacaacttc | 2220 |
| QY | 2287 | CAATAGTGCATTCCGGAAACATCTCAAGTCCAGACAGTACTGTTGTATTACCCACAGTT | 2346 |
| Db | 2221 | caatagtgctataccggaataatctcaagctccaaagacagcttcgtttgtatacccaacagtt | 2280 |
| QY | 2347 | ACAGTACCTGTTGACTGTGATGATGATGATCTTTCATGAAGAAGGCGCTGATTACGGAAAG | 2406 |
| Db | 2281 | acagtaacctgctacgtgatgaagtgatcttcatalgaagaagggctgattatcagaaag | 2340 |
| QY | 2407 | AGGCACCATATGAGAACTGATGAATTTAAATGCTGACTTGTACTACATTTTAAATAACT | 2466 |
| Db | 2341 | aggcacccatgaagaactgatgaattaaattgctactatgctacacattttaaataacct | 2400 |
| QY | 2467 | GTTGCTGGAGAGACACCGGCACTGTGAGATCAATTTCAAAAAAGAAACAGATGTTTCA | 2526 |

```
||||| 2401 gttcctgagagacacgcgcatgtgataatcaaaaaagaaacagtggtccaca 2460
||||| 2527 GAAGAAAGCACAAAGACAGGCTCTAAACAGATCAATTAAGAAAGAAAAAGCAGTAAA 2586
||||| 2461 gaagaatcaacaagaacaaaggttcttaaaacagatcaataaagaagaaagacgtataa 2520
||||| 2587 GCCAGAGGAAGGACGCTTGTGACGTGGAAGAGAAAGGAGGAGGCTTACGTCCTGCTC 2646
||||| 2647 AGTATATGTTCTTACATCCAGGCTGCTGGGGGCCCCCTGGCATTCCTGGTTATATGCG 2706
||||| 2581 agtatagtgtctacaatccagcgtctggtgggcccccttgcatctcgttatatagtgc 2640
||||| 2707 CCTTTTCATGCTGATGATGAGCAGACCCGCTTACGACCTGCTGATGATGATGATGAT 2766
||||| 2641 cctttcactgtcgaatgtaagcagacaccccttcagcaccctggtgtgattactgcat 2700
||||| 2767 CAAGCAAGGAAGCGGGAAACACCACTGTGACTGAGGGAAGAGACCTGGTGTGACAG 2826
||||| 2701 caagcaagaagcgggaacacacactgtactgagggagagactcgtgtgagtgcag 2760
||||| 2827 CATGAAGACATCTCTATATGACGATCTATGACGATCTTACGCCCTTCCATGCGCAGT 2886
||||| 2761 catgaagacaatccctcaatgacgactatgcaagacatcaagccctccatgagcagt 2820
||||| 2887 CATCTGTATCTGAAAGGACATTCGAGAGTGTGCTTGTCAAGGGCAGCTGCGAGCTTC 2946
||||| 2821 catgtgtactgtgaagcactatcgagggagtgtctcttgtaagggcaagcgtcgagcttc 2880
||||| 2947 CTCGCCGCTGATGACGAGCTTTTCCGAGAGATCTCTTGAAAGCCCTATGAAGTTTGA 3006
||||| 2881 ctcccggtctgcatgacgagcttctccgaagatcccttcgaagccctatgaagcttcttga 2940
||||| 3007 CACGACCCCAAGGAGGAGATTCACAGGTTTCCAAAGACATGATGATGATGATGAGT 3066
||||| 2941 cagaccccccaaggaagatctccaagaagcttccaagaacagatgagatgagcagt 3000
||||| 3067 GCGGCTGCCCTTCCAGGCGCGAGATGTTTCATCCAGAACGTTTCCGATGCTTCTGTGT 3126
||||| 3001 ggcgctgcgctccagcgcgagatgtctatccagaagcttaccgtgtctctctgtgct 3060
||||| 3127 GGGATATATCCAGAGAGCTTCCCGTGTCTTGTGACATGGGGGCCCTTGTATCTCT 3186
||||| 3061 gggatgtatccagagagcttcccggtgtctctgtgcaagtggggccccctgtatctct 3120
||||| 3187 CTTTTCATGCTCGACATGTTCTCCAGGCTCTGATTTGGGAGCTGAAGCGCTGACAA 3246
||||| 3121 ctcttcagctcgtcaaatgtctctccagagcttcgagctcggaagcgtctcgtgacaa 3180
||||| 3247 TATGACGACATCACCCTTCTCTCCACATCAGTCCAGATACAGGGCCTTGCCACCAT 3306
||||| 3181 tatcagcagtcacacttctctccacatacgtccagatacagggccttgccacacat 3240
||||| 3307 CCAGGCGCTACATTAAGGGCAGAGAGTTCCTGCACAGATACCAAGAGCTGTGATGA 3366
||||| 3241 ccaagcctacaataaagggcagaggttctctgacagatacaagagcgctcgtgagacaa 3300
||||| 3367 CCAAGCTCTCTTTTCTTTTGTAGCTGTGCGATGGGCTGGGCTGTGCGGCTGGACCT 3426
||||| 3301 ccaagctcctcttcttctgttaacgtcgtcgaagtggtgtgctgtggtcgtgagccct 3360
||||| 3427 CATGACGATCGCCCTCATACACACACAGGGGCTGATGATGCTTTATGCAAGGGCAGAT 3486
||||| 3361 catcagcactcgccctcatcacacacacagggcgtgactgcttctatgcacggccagat 3420
||||| 3487 TCCCCAGACCTATGCGGGGTCTCGCATCTCTTATGCTGTCCAGTTAAAGGGGCTTCCA 3546
||||| 3421 tccccagcctatctggtgtctcgccatctctatgctgtcagttaaagggtcgtgttcca 3480
||||| 3547 GTTTACGCTGACGCTGAGCTGAGACGAAGCTGATTCACCTGGTGGAGAGATCAA 3606
|||||
```

```
Db 3481 gtttaaggtcagactgcatctgagacagaagctcogatctacccctcggtgtgagagatcaa 3540
QY 3607 TCACATCAATTAAGACTGTCTCTTGGAAACACTCCGCAAGATTAAGAAACAAGGCTCCCTC 3666
Db 3541 tcaactacaatgaagctctgtctcttggaagcactcgtccagaattaaagaacaggtccctc 3600
QY 3667 CCCTGACTGGCCCCCAGAGGAGGAGAGTACCTTTGAGAACGACAGATGAGACTACCGAGA 3726
Db 3601 ccttgactcgtgccccagagggagaggtgtaaccttgaagacgcaagatgagaggtacagaa 3660
QY 3727 AAACCTCCCTCGCTCTAAAGAAAGTATCCCTTCACGATCCAAACCTTAAGAGAGATTGG 3786
Db 3661 aaacctccctctgtctcacaagaagatcccttcacgaatcaacctaaagagaagattgg 3720
QY 3787 CATGTGGGGGGAAGAGATCAGGGAAAGTCTCGTGGGGATGGGCCCTTCCGCTTGGT 3846
Db 3721 catgtgtggcggacaggaatacaggaagctcctcggtgggtggcccccttcgtctggt 3780
QY 3847 GGAGTTATCTGGAGGCTCATCAAGATTGATGAGTGAATCACTGATATTGGCTTGC 3906
Db 3781 ggaatlatctgaggtctgcatcaagatgatagtgagatgaaatcaagtataatgtgccttc 3840
QY 3907 CGACCTCCGAAGCAAACTCTGTATCATCTCTCAAGAGCGCGTGTGTCAGTGGCACTGT 3966
Db 3841 cgacctccgaagcaaacctctcatcaatccctcaagagcgggtgtgtctcagtgagcactgt 3900
QY 3967 CAGATCAAAATTTGACCCCTTCAACCACTACACTGAAAGCAGATTGGGATGCCCTTGA 4026
Db 3901 cagatcaaatcttgagcccccttcaaccagttacacgtgaagcagatttggtgtcccttga 3960
QY 4027 GAGGACACACATGAAAGAAATGTTGCTCAGCTTACCTCTGAAACTTGAATCTGAAGTAT 4086
Db 3961 gaggacacacatgaaagaatgatatgtcgaactacctctgaaacttgaaatctgaaagtat 4020
QY 4087 GGAGATGGGATTAATCTCTCAGTGGGGGAACGGCAGCTCTTGTGACAGCTTAAGCCCT 4146
Db 4021 gggagaatgggataaactctcaagtgggggaagcagagctctgtgataagtaagccct 4080
QY 4147 GCTCCGCACTGTAAAGATTGTGATTTTATGATGAAGCCACAGCTGCCATGACACAGAGAC 4206
Db 4081 gctccgcacagttaagatcttgatttagaagcccaagctgcacatgagacacagaagac 4140
QY 4207 AGACTTATGATTCOAAGGACATCCGGAAGCATTTGACAGCTGATACATGATGACCAT 4266
Db 4141 agaatctatgtatccaagagacacacccgaagacatctgcgaacgttacaatgctgagacat 4200
QY 4267 TGCCCATGCCCTGACACACAGGTTTACGCTCCGATAGATTAATGATGCTGAGCCCAAGGACA 4326
Db 4201 tgcacatcgctcgtcaacaggtctcaggtccgataagattatgtgtcgtgcacaggagca 4260
QY 4327 GGTGTGAGATTTGACACCCCATGCGTCTTGTGTCCACAGACAGTTCCGATTTATGC 4386
Db 4261 ggtgtgtgaattgacaccccatcgctcctctgtcacaagacaggttcccgattctatgc 4320
QY 4387 CATGTTGCTGTGAGAGCAAGGTGCTGTCAAGGGCTGACGCTCCCTGTTGACGA 4446
Db 4321 catgtctgtcgtcgaagaaagtcgctcgaagagcgacccctcctgttgagaa 4380
QY 4447 AGTCTCTTCTCTTAAAGACATTTGCCATTCCTCGCTGGGGGGGGCCCCCTTCATGCGCTCC 4506
Db 4381 agtctcttctcttaagacatgcmgkymbmbkctcgtgggggcccccttcatcgtcgtcc 4440
QY 4507 TCTTACGGAACCTTGGCTTCTGTGATTTTATCTTGTCCACAGACAGTTCCGATTTGCTT 4566
Db 4441 tctctacgaaacctgtccttctcgatttctctcgcacaagcaggtccgattgtgctt 4500
QY 4567 GTGTGTTTCATTTTAAAGGAGATGATTTTGAATTTGATTTATTTATTCATTTATCATGT 4626
Db 4501 gtgtgttcaactttaaaggagatcaatcttgatatactgtatcaatcaatcaatcaatc 4560
QY 4627 AAACAAATTTAGTTTGTCTTAAATGCACTCTAAAGGTTCAAGGAAACGTTATTTAT 4686
Db 4561 aaacaaatttagtttctgtctcaatctcaatctcaatctcaatctcaatctcaatctat 4620
```

| | | | |
|--------|---|--|------|
| QY | 4687 | AATGTATCAGGCGCTTAAATGAAGCTTTAAGCGTAGTCATATCATATAATTTC | 4746 |
| Db | 4621 | aattgatacagaggccataaaagaagctttacaagtgtagcatatcataataatct | 4680 |
| QY | 4747 | GTCATACCCATATTTTACAGTAAGAATGTAAAGCTGTATTTTATATATAAATAGCAC | 4806 |
| Db | 4681 | gtacatatgccatatttaccagtgaanaatgtaagcgctgtattttatataataaagcac | 4740 |
| QY | 4807 | TGTGCTAAAAAAAAAAAAAAAAAAAAAGCGGCCGC | 4847 |
| Db | 4741 | tgtgctaataaaaaaaaaaaaaaaaagcgccgc | 4781 |
| RESULT | 6 | | |
| ID | AAT94054 | standard; cDNA; 1761 BP. | |
| XX | AAT94054; | | |
| DT | 01-APR-1998 | (first entry) | |
| DE | Human multidrug resistance-associated protein 5 (MRP5) cDNA sequence 1. | | |
| KX | Canalicular multispecific organic anion transporter protein; | | |
| KM | cMOAT protein; ATP-binding cassette transporter family; ABC transporter; | | |
| KW | hepatobiliary excretion; multidrug resistance-associated protein; MRP5; | | |
| KM | cMOAT protein activity; multidrug resistance-related protein; MDR-1; | | |
| XX | Dubin-Johnson disease; Rotor disease; ss. | | |
| XX | Homo sapiens. | | |
| PN | WO9731111-A2. | | |
| PD | 28-AUG-1997. | | |
| XX | 21-FEB-1997; | 97WO-NL00079. | |
| PR | 22-FEB-1996; | 96EP-0200460. | |
| PA | (MEDT-) ACAD MEDISCH CENT AMSTERDAM. | | |
| PA | (HETN-) HET NEDERLANDS KANKER INST. | | |
| PI | (INTR-) INTROGENE BV. | | |
| PI | Borst P, Bosma PJ, Evers R, Oude Elferink RPJ; | | |
| PS | Paulusma CC; | | |
| PT | WPJ; 1997-435163/40. | | |
| PT | DNA encoding human and rat canalicular multispecific organic anion | | |
| PT | transporter proteins - useful for diagnosis and treatment of | | |
| PT | Dubin-Johnson disease and Rotor disease | | |
| XX | | | |
| XX | Disclosure; Fig 19A; 106pp; English. | | |
| XX | The present cDNA sequence encodes a human multidrug resistance-associated | | |
| CC | protein 5 (MRP5). The MRP5 gene has been localised to chromosome 3. The | | |
| CC | MRP5 protein is a member of the ATP-binding cassette (ABC) transporter | | |
| CC | family of anorganic anion transporters. MRP5 is a 4 domain protein, with | | |
| CC | 2 ATP-binding domains, and 2 domains with transmembrane regions. The | | |
| CC | protein is a homologue of MRPl, which is involved in ATP-dependent | | |
| CC | transport of glutathione conjugates such as dinitrophenyl glutathione. | | |
| CC | These substrates are also transported by a novel canalicular | | |
| CC | multipspecific organic anion transporter (cMOAT) protein. The ATP | | |
| CC | dependent cMOAT transporter system mediates hepatobiliary excretion in | | |
| CC | the liver. cMOAT may be a liver-specific homologue of multidrug | | |
| CC | resistance-associated protein. The nucleic acids are used to provide | | |
| CC | cells with cMOAT protein activity. cMOAT protein activity in cells can | | |
| CC | be enhanced by increasing the level of glutathione, glutunone and/or | | |
| CC | sulphate. Antisense constructs, especially derived from another multidrug | | |
| CC | resistance (MDR)-related protein, e.g. MDR-1, to the nucleic acids and | | |
| CC | vectors can be used to decrease the level of cMOAT in a cell. The nucleic | | |
| CC | acids and proteins can be used especially in diagnosis of Dubin-Johnson | | |

CC disease, Rotor disease or another disease involving cMOAT. The cMOAT gene
CC may also be used as a selectable marker gene.
XX
S0 Sequence 1761 BP; 427 A; 448 C; 469 G; 413 T; 4 other;

| | | | | |
|----------------------------|--------|---------------|--------|-------------------|
| Query Match | 34.5% | Score 1674.4; | DB 18; | Length 1761; |
| Best Local Similarity | 99.5%; | Pred. No. 0; | | |
| Matches 1722; Conservative | 0; | Mismatches | 1; | Indels 7; Gaps 4; |

Oy 1969 CATTGCAATCAGTGGAACTTCGCTTATGTGCCCCACAGCGCTTGATCCCTCAATGCTAC 2028
|||||
|||||
Db 1 cattgcaatcagtgaaccttcgtctattgcggccagcagg-ctgatcatcataatgctaac 59

0y 2029 TCTGAGAGACAACATCCCTGTTGGGAGGAATATGATGAAGAAAGATCAACTCTGTGCT 2088
 db 60 TCTGAGAGAGCAACTCTATTTGAGGAGGATATGATGAAGAAAGATCAACTCTATCT 119

2089 GAACAGCTGCTGCCTGAGGCTGACCTGCGCATTCCTCCACAGCGACCTGACGGAGAT 2148

Db 120 gaacagcigtctgcctgagagcctgacctgtgcattcttcccaagcagcgacctgacgagat 1/9

Q7 2149 TGGAGAGCGCAGGAGCCAACTGAGCGGTGGGCAGCGCCAGAGGATCAGCCTTGCCTGGGC 2208

| | | | |
|----|------|---|------|
| Db | 180 | tggagagcgagagccaacctgagcggctgggcagcgcagagagatcagcctctgcccggc | 239 |
| Qy | 2209 | cttctatctagtgacagagcattctacatccttgacgacgacccccctcagtcgctttagatgccca | 2268 |

2269 TCGGGGCAACACATCTTCATATAGGCTATTCGGAAACATCTCAAGCTCAAGACAGTTCT 2338
Db 240 ctgtatagtgacagagcatctacatccctgacgacccccctagtccttagatgccca 299

Db 300 tgtgggcaaccacatctcaatagtgtctatccggaacaatctcaagtcacaagactct 359

Qy 2329 GTTT---GTTACCAACAAGTATACAGACCCTGTACTGTGATGAAGATCTTCATGAA 2385
||| ||||||||||||||||||||||||||||||||||||||||
Db 360 gtttmggttacccacagttacagtaacctggttgactgtgatgaagtatcttcataaa 419

Qy 2386 AGAGGGCTGATTACGGAAAGAGGCACCCATGAGGAACGTATGAAATTTAAATGCTGACTA 2445
 |||||
 Db 420 agaggcgctattacggaagagcgaccccatgaggaactgtacgaatttaattggtgacta 479

Oy 2446 TGCACCATTTTAAATTAACCTGTTGCTGGGAGAGACACCGCAGTTTGATCAATTCCAAA 2505

Dh 480 TGTATACCATTTTTATAAAGCTATTGATGAAGAGAAAAAGCCGCAATTAGATGAATTCGAA 539

2506 AAAGGAAACCGTGGTTCACGAGAGAGTCAACAGACAAAGGTCCTAAACAGATCAAT 2555

Db 540 aaaggaacaccagtgtctcacgaagaagtcacaaagcaagggctcctaataacagatcagc 599

QY 256 AAAGGAGGAAAAAGCAGTAAAGCCACAGGAGGAGGCGACCTTGTCCAGCTGGAAACAGAAAG 26259

db 600 aaagaaagaaaacagtaaacacagaggaagagcagcttgcgcagctgcgaagagaag 659

QY 2626 GCAGGGTTCAGTGCCTGGTCAGTATATGGTCTACATCCAGGCTGCTGGGGCCCTT 2685

2656 GGGCA-TTTCCTGGTTATTATATGGCCCTCTTTTCAGG-C-CTGAATGTAAGGACACCCGCGCTTCAG 2742
 Db 660 gcaaggttcagtcgccctgctcagtatatgtgtctctacatccaggtctgtggtggccctt 719

Db 720 ggcaattcctgtatattatgcccctttcatgagcctcgaatgagcagcagcctcag 779

97 2743 CACCTGGTGTGAGTACACGGATCAAGCAAGCAAGCGGGAAACACCACTGTGACTCGAAG 2802
 |||||
 Db 780 cactctgtgttgagttacttgcataagcaaggaagcgcgggaacacacactgtgactcgag 839

QY 2803 GAACGAGACCTCGTGAGTGCACGACGATGAAGGACCAATCTCATATGCAAGTACTATGCGAG 2862
|||||
Db 840 gaacgagacctcggtgtgagtacacgacatgaagacaatcctcatatgctgactactatgccag 899

Qy 2863 CACTACGCCCTCTCCATGGCAGTATGCTGATCCGAAAGCATTCGAGGAGTGTCTT 29222

```

QY 2923 TGTCAAGGCGACGCTGCGACGCTTCTCCCGGTGATGACGAGCTTTTCCGAGAGATCCT 2982
    |||
Db 960 tgcgaagcgacgctcgagctctcccgctgcatgacgagctctccgaagatcct 1019
    |||
QY 2983 TCGAGCCCTTGAAGTTTGTGACACGACCCCGACAGGAGATTTCAACAGCTTTTC 3042
    |||
Db 1020 tcgaagccctatgaagttttttgacagaccccccacagagagatctcacaagtttcc 1079
    |||
QY 3043 CAAGACATGATGAAGTTGACGTGCGGCTGCGCTTCCAGGCCGAGATGTTTCACGAA 3102
    |||
Db 1080 caaagaacatgtagagtgacgtgcgctgcgtccagagcgagatgltcatccagaa 1139
    |||
QY 3103 CGTTATCCTGGGTCTTCTGTGTGGGAATGATGCGACGAGAGTCTTCCGTGTTCTTGT 3162
    |||
Db 1140 cgttacctgtgtctctctgtgtggaatgacgcagagagcttcccggttctctgt 1199
    |||
QY 3163 GGCAGTGGGGCCCCGTGATCATCTCTTTTCACTGCTGACATGTCATCAGAGGTCTGTAT 3222
    |||
Db 1200 ggcagtgaggccctgttcattccctcttcagctcgtcacatgtgtccagaggtctgtat 1259
    |||
QY 3223 TCGGAGCGTGAAGCGTCTGACATATCAGCAGTCACTTCTCTCCACATCAGCTC 3282
    |||
Db 1260 tcggagagctgaagcgctctggaacataatcagcagtcacacttccctccacatacgtc 1319
    |||
QY 3283 CAGCATACAGGSGCTTGGCCACCATCCAGGCTTACAAATAAGGGCAGAGATTCTGCACAG 3342
    |||
Db 1320 cagcatacagggcctctgcacacatccacgacctacaataaaggcagagttctctgcacag 1379
    |||
QY 3343 ATACAGAGAGCGTGTGATGACACCAAGCTCTTTTGTGTAGTGTGCGACGTGCG 3402
    |||
Db 1380 atacagagagcgctcgtgagacacaaagctctcttcttctgttcaagtcgtcgatgctg 1439
    |||
QY 3403 GTGGCTGCGCTGTGCGCTGACCTCATCAGATCGCCCTCATCACACACGCGGCTGAT 3462
    |||
Db 1440 gtggctgctgtgctgctgagacctatcaagctgcgcctccacacacacagggctgct 1499
    |||
QY 3463 GATCGTTCTTATGACACGGGACAGATTCCCGACGCTATGCGGGTCTGCGCATCTTATGTC 3522
    |||
Db 1500 gatcgttctttagcagcaggtcagatcccccagcctcaltgctgctcgtcatctctatgc 1559
    |||
QY 3523 TGTCCAGTTAAAGGGCGTGTTCAGTTTACGTTTACGATCGAGATCGATCGAGACGAGCTG 3582
    |||
Db 1560 tgcctagttaaagggcgctgtcccaagttcaaggtcagacgtgcagatctgacacagaagctgc 1619
    |||
QY 3583 ATTCACTCTGCTGAGAGAGATCAATCACTACATTAAGACTGTGCTTGAAGCACTGTC 3642
    |||
Db 1620 attcaactcctgctgagagatcaatcaactaactaagactctgtctcttggaaagcactgc 1679
    |||
QY 3643 CAGAAATTAAAGAACAGGCTCCCTCCCTGACTGCGCCCGACGAGGAGAGAG 3692
    |||
Db 1680 cagaatttaagaacaaggtcctccctccctgactgctcccaagagagagag 1729
    |||

```

```

XX 28-AUG-1997.
PD 21-FEB-1997; 97WO-NL00079.
XX 22-FEB-1996; 96EP-0200460.
XX (MEDI-) ACAD MEDISCH CENT AMSTERDAM.
XX (HETN-) HET NEDERLANDS KANKER INST.
XX (IMNR-) INTROGENE BV.
PI Borst P, Bosma PJ, Evers R, Oude Elferink RPJ;
PI Paulusma CC;
XX WPI: 1997-435163/40.
DR DNA encoding human and rat canalicular multispecific organic anion
PT transporter proteins - useful for diagnosis and treatment of
PT Dublin-Johnson disease and Rotor disease
XX PS Disclosure; Fig 19B; 106pp; English.
XX PS
CC The present cDNA sequence encodes a human multidrug resistance-associated
CC protein 5 (MRP5). The MRP5 gene has been localised to chromosome 3. The
CC MRP5 protein is a member of the ATP-binding cassette (ABC) transporter
CC family of anorganic anion transporters. MRP5 is a 4 domain protein, with
CC 2 ATP-binding domains, and 2 domains with transmembrane regions. The
CC protein is a homologue of MRP1, which is involved in ATP-dependent
CC transport of glutathione conjugates such as diltrophenyl glutathione.
CC These substrates are also transported by a novel canalicular
CC multispecific organic anion transporter (cMOAT) protein. The ATP
CC dependent cMOAT transporter system mediates hepatobiliary excretion in
CC the liver. cMOAT may be a liver-specific homologue of multidrug
CC resistance-associated protein. The nucleic acids are used to provide
CC cells with cMOAT protein activity. cMOAT protein activity in cells can
CC be enhanced by increasing the level of glutathione, glucuronide and/or
CC sulphate. Antisense constructs, especially derived from another multidrug
CC resistance (MDR)-related protein, e.g. MDR-1, to the nucleic acids and
CC vectors can be used to decrease the level of cMOAT in a cell. The nucleic
CC acids and proteins can be used especially in diagnosis of Dublin-Johnson
CC disease, Rotor disease or another disease involving cMOAT. The cMOAT gene
CC may also be used as a selectable marker gene.
SQ Sequence 2167 BP; 494 A; 511 C; 511 G; 626 T; 25 other;

```

```

Query Match 21.7%; Score 1054.2; DB 18; Length 2167;
Best Local Similarity 96.0%; Pred. No. 9.4e-262;
Matches 1094; Conservative 7; Mismatches 36; Indels 3; Gaps 2;

```

```

QY 3682 GGAGGAGAGGTGACCTTTGAGACGACGAGATGAGTACGAGAAACCTCCTCTGCT 3741
    |||
Db 29 gsaagkkracntcaltcagaaagacraaagatgtgtaagsagaacttccttltat 88
    |||
QY 3742 CCTAA--GAAGTATCCTTCACGATCAACCTTAAGAGAGATGAGATGTTGGGGGG 3799
    |||
Db 89 cctaaagaagatgctcctcagatcaaacctaaagaagaatgtgactgtggtggcg 148
    |||
QY 3800 ACAGGATCAGGAGAGTCTGCTGCGGATGCGCTTCCGCTGCTGAGTATTAATCTGA 3859
    |||
Db 149 acagatcagggaaagctcctcgtggtggtggtggtccttcctcgtctgtgagtactg 208
    |||
QY 3860 GCGTCATCAGATTGATGAGTGAATCACTGATATTTGGCTTCCGACCTCCGAAGC 3919
    |||
Db 209 gctgcatcaaatgattgagtgagtaacatcagatgatttggtccttcgacccgaagc 268
    |||
QY 3920 AAACGCTCATATTCCTCAAGAGCGGCTGCTCAGTGGACATGTCAGATCAAAATTG 3979
    |||
Db 269 aaactctcatatccctcaagagccggtgctgctcagtgacgacgtcaacaattg 328
    |||
QY 3980 GACCCCTTCAACAGTACACTGAAGACAGATTTGGATGCCCTTGAGAGACACACATG 4039
    |||
Db 329 gacccttcaaacaggtacactgaagacagagatttggtggtcccttgagagacacatg 388
    |||

```



```

OY 4040 AAAGATGATTTGCTACGCTACCTGGAACCTTGATCTGAATGATGAGATGGGAT 4099
    |||
DB 389 aagaatgatattgctcaccctacacctggaacttgaaatctgaaagtatgaaatgagat 448
OY 4100 AACTTCTAGTGGGGAAGCGGAGCTCTGTGTCATAGCTAGAGCCCTGCTCCGCACTGT 4159
    |||
DB 449 aacttcgaagtgagggaacgagctctgtgcatagctagagccctgtccgcacagt 508
OY 4160 AAGATTCGATTTTATGATGAAGCCACAGCTGCGCATGACACACAGACAGACTTATGATT 4219
    |||
DB 509 aagattcgtattttagatgaagccacagctgcatgacacagacagacattatgatt 568
OY 4220 CAAGAGACCATCCGAGAGCAATTCGACACTGATACATGCTACATGCCATGCCGCTG 4279
    |||
DB 569 caagagaccatccgagagacattgcaagactgtaccatgtgacacattgcccacatgctg 628
OY 4280 CACAGGTTCTTAGCTCCGATAGATTAAGTGTGCTGCGCCAGAGGACAGTGGTGAATT 4339
    |||
DB 629 cacaaggtctcagctcgcgatagatagtgtgctgcccagagagagtggtgagatt 688
OY 4340 GACAGCCCATCGGCTCTCTGTGCAACGACAGATCCCGATTCTATGCCATGTTGCTGCT 4399
    |||
DB 689 gacacccatcgctctctctgtcacaagacagctcccgattctatgccaagtgtgctgct 748
OY 4400 GCAGAGAACAAAGTGCCTGTCAAGGGCTGACTCCCTCCCTGTGAGAGAACTCTTTCTT 4459
    |||
DB 749 gcaagagaaacaaagtctgtcacaagctgactccctcctgtgacgaagctctcttctt 808
OY 4460 TTAGACATTTGCTATCCCTGCTGGGGGGGGGGCCCTTCATCGCGTCCCTACCGAAAC 4519
    |||
DB 809 tagagcatgtccatccctgctgctgaggggccccc- tcatgctgctctccacgaacc 867
OY 4520 TTTGCTTTCTCGATTTTATCTTTGCGACAGAGATCCGGATTGGCTGTGTGTTCACTT 4579
    |||
DB 868 ttgctcttcgacttattatcttcgcacagcagctccgattgctgtgttccactt 927
OY 4580 TTAGGAGAGCTATTTTGTATTTGATTGATTGATTATTCATATTCATGTAACAAATTTG 4639
    |||
DB 928 ttaaggagagctaatatttgattatttatttatttccatattcattgtaacaaatttag 987
OY 4640 TTTTGTGTTCTTAAATGCACTCTTAAAGGTTTCAGGAACCGTTATTATTAATTTATCAG 4699
    |||
DB 988 ttttcttcttaattgtcactctcaaaagttcaggaaccgttatcataatgtatcagag 1047
OY 4700 GCCATTAATGAAGCTTTATACGTGTAGCTATATCTATATTAATTTCTGTACATAGCCTAT 4759
    |||
DB 1048 gccataatgaagcttatacgtlgtactatatactataatctgtacaatagcctat 1107
OY 4760 AATTACAGTGAATGTAAAGCTTTATTTTATTTAAATTAAGCACTGTGCTAAAAAAA 4819
    |||
DB 1108 attacagtgaaatgtlaagctgttatttattataataaataagacactgtgctaataca 1167

```

```

FT FT /product= "human MRP variant"
FT 2249
FT variation
FT /tag= b
FT /note= "a change from wild-type T to C"
FT variation
FT /tag= c
FT /note= "a change from wild-type C to G"
FT variation
FT /tag= d
FT /note= "a change from wild-type G to C"
FT US5766880-A.
FT 16-JUN-1998.
FT 05-JUN-1995; 9505-0463092.
FT 05-JUN-1995; 9505-0463092.
FT 05-JUN-1995; 9505-0463092.
FT 27-OCT-1992; 92US-0966923.
FT 08-MAR-1993; 93US-0029340.
FT 26-OCT-1993; 93US-0141893.
FT 20-MAR-1995; 95US-0407207.
FT (TOOH ) UNITV QUEENS KINGSTON.
FT PA
FT Cole SP, Deeley RG;
FT WPI; 1998-361687/31.
FT P-PSDB; AAM57486.
FT DNA encoding protein associated with multi-drug resistance - useful
FT for as probe for identifying multi-drug resistant tumour cells
FT PT
FT Claim 12; Columns 67-78; 82pp; English.
FT XX
FT PS
FT CC This cDNA encodes a variant of the human multidrug resistance-associated
FT CC protein (MRP). This natural variant is a leishmania P-glycoprotein related
FT CC molecule ltrpPa (lcl/PgPa). The human and murine MRP nucleic acid
FT CC molecules can be used as probes for identifying multidrug resistant
FT CC tumour cells by hybridisation to mRNA from tumour cells. The antisense
FT CC nucleic acid can be used to reverse multidrug resistance (MDR). A
FT CC recombinant expression vector containing the MRP nucleic acid molecules
FT CC operatively linked to at least one regulatory sequence can be used to
FT CC transform a host cell to produce a recombinant MDR-associated protein.
FT CC
FT SQ
FT Sequence 5011 BP; 1064 A; 1498 C; 1394 G; 1055 T; 0 other;

```

```

Query Match 10.5%; Score 511.2; DB 19; Length 5011;
Best Local Similarity 52.5%; Pred. No. 2e-121;
Matches 1367; Conservative 0; Mismatches 1178; Indels 59; Gaps 9;

```

```

OY 1811 CACATTCACCTGGGCGACCTGCGCTTACAGAGGACACTGCACAGCATCGATCGAGATC 1870
    |||
DB 2142 cacattcactctggcca--ggagcgacccctccacactgaatgataccttccatc 2199
OY 1871 CAAGAGGGTAAACTGTTGGAATTCGCGCATGTGGGAAGTGGAAAACTCTCTCATTT 1930
    |||
DB 2200 ccggaagtgcttctgtgctgctgctgctgctgctgctgctgctgctgctgctgct 2259
OY 1931 TCAGGCATTTTAGGCCAGATGAGCGCTTACAGAGGCGCAGCATGAGTGAACCTTC 1990
    |||
DB 2260 tcagccctctgtctgagatctggaacaaagtggagggcagctgctatcaaggtccctg 2319
OY 1991 GCTTATGAGCCGACGAGCGCTGATCTCAATGCTACTGTGACACACAATCTGTTT 2050
    |||
DB 2320 gctatgtgcacaagcagcgtgattcagaatgtattctctccgaaaaacatcccttlt 2379
OY 2051 GGGAAAGATATGATGACAAGAAAGATPACACTCTGTCGTAACAAGCTGCTGAGGCGCT 2110
    |||
DB 2380 ggatgtcagctgtgaggaacacatactacagctcgtatatacagcgtgtccctccca 2439
OY 2111 GACTGCGCATTTCTTCCGACGACGACTGACGAGATGAGAGAGGAGGACCAACTGT 2170

```

[illegible]

| | | | |
|---|------|---|------|
| D | 3520 | ctctgtgccaagcccaatcgccgcacatcatcatccgccctctgtccatctactcttc | 3579 |
| Q | 3197 | CTGCACATTTGTTCTCCAGGGTCTGATTCCGGAGCTGAAAGCTCTGGACAAATTACAGCA | 3256 |
| D | 3580 | gtccaaagattctctacgtgtctctccgcgcagcttaagcgcctcagtcgacgcgc | 3639 |
| Q | 3257 | TCACCTTTCCCTCCCAATCAGTCCAGCATTAAGGGCTTGCCACCATCCAGCCATAC | 3316 |
| D | 3640 | tcccgggtctattccattccattcaagaaacctctgtggtggatcagcgtactcttgagcctt | 3699 |
| Q | 3317 | AATTAAGGGGAGAGTTTCTGCACAAATACAGAGACCTCTGGATTGACAACTCACTCT | 3376 |
| D | 3700 | gagaaagcaagaaagcgtcttcaaccaagaagtgtacctaaggttgaaagaaagaaagacc | 3759 |
| Q | 3377 | TTTTTTTTTTTACGTGTGCGATGCGGTGGCTGTGTCGGCTGTGACCTCATACGATC | 3436 |
| D | 3760 | tattaccocagacatcgtgcgcaacaagatgtgtccgtgtgcgtgtgagtgtgtgtgggc | 3819 |
| Q | 3437 | GCCCTCATCACACACACGGGGGTGATGATCGTTCTATATGACAGGGGCAAGTTCCCCAC | 3496 |
| D | 3820 | tgcacgtctctgtttgtgtgctccgtttgtcggtatctcagaagcaagcctcagttgtgc | 3879 |
| Q | 3497 | TATCGGGGTCTCGGCATCTCTTATATGCTGTACATTAACGGGGCTGTCCAGTTATCGT | 3556 |
| D | 3880 | ttgtgttggtcctcagatgtcttactcatcttgcaggtacccaagttactgaactgtgtgt | 3939 |
| Q | 3557 | AGACTGGCATCTGAGACGAAGCTCGATTCCCTCGGTGGAGAGATCAATCACTACAT | 3616 |
| D | 3940 | cgagtgtcatctgaattgaaaccacaacatctgtgcccgtgtgagagctcacaagattcca | 3999 |
| Q | 3617 | AAGACTGTGTCCTTGGAAAGCACTGCCAAGATTAAACAAGGCTCCCTCCCTACCTGG | 3676 |
| D | 4000 | gagactctagaaa--gagagcgccctgtgcacatccaagaaagacgtctccgcagcagcttg | 4056 |
| Q | 3677 | CCCCAGGGAGAGAGGAGACTTGTGAAAGCGCAGATAGAGGACCGAATAAACCTCCCT | 3736 |
| D | 4057 | ccccaggatgtgcgagtgaaatcccgaaactactctcgtgcctacccaagaaagacctgac | 4116 |
| Q | 3737 | CTCGTCTTAAGAAAGTATCTTACAGATCAAACTTAAGAGAAAGTTGGCACTTGTGGGG | 3796 |
| D | 4117 | ttcgtctcagagcaatcaatgtcacaatcgaatcgaatggggaagaaagtcgtacatgtggg | 4176 |
| Q | 3797 | CGGACAGGATCAGGGAAGTCTCGCTGGGATGGCCCTTTCCTGCTGTGGATTAATCT | 3856 |
| D | 4177 | cysgcgggaactgtgaaagtcgtccctgtgacccttgggtatttcggtatcaacgaagtcgc | 4236 |
| Q | 3857 | GGAGGCTGCATCAAGATTGATGAGAGAGAAATCAGATTAATGGGCTTCCGCACTCCGA | 3916 |
| D | 4237 | gaaggaagagatcatcatcgaatctgtgtatcatcaatctgcgaagatctgcgttcgacaccc | 4296 |
| Q | 3917 | AGCAAACTCTCTATCATCTCTCAAGACCGGTGCTCTTCAATGTCGACTGTGCATCAAT | 3976 |
| D | 4297 | ttcaagatacaatcatcctccccaagaaacctgtttgttttgcgttccctccgaatgac | 4356 |
| Q | 3977 | TTGGACCCCTTCAACAGTACACTGAAGACAGATTGGGAATGCCCTGGAGAGGACACAC | 4036 |
| D | 4357 | ctgagccattcagccagtaactcgtatgaagaagtcgtgaagctccctgtgaagctgtgccac | 4416 |
| Q | 4037 | ATGAAGAAATGTATGCTGTACGCTACCTCTGAAACTTGAACTGTGAAGTAATGG | 4096 |
| D | 4417 | cttgaagaactctgtgtacagccctctctgcacaagctcagaccatgaaatgtgcagaaggg | 4476 |
| Q | 4097 | GATTAACCTTCTACAGGGGGGAAAGGAGCTTTGTGATAGCTAGAGACCCTGTCCGCCAC | 4156 |
| D | 4477 | gagaaactcagttctgtgcagcagcgcagcttgtgtgtcctgaccccggtccctgtctgagaa | 4536 |
| Q | 4157 | TGTAAGATTCTGATTTTATGATGAAGACACAGCTGCACATGGACACAGACACTTAATG | 4216 |
| D | 4537 | acgaagaatccttgtgtgtgtgagcgaagccgctgtgacctgtgaaaggaagacatc | 4596 |
| Q | 4217 | ATTTAAGACATCCGAGAAAGATTTCAGAGCTGTACATGCTATACATTGCCATATGC | 4276 |
| D | 4597 | atccagtcacacatctcggaaacagttctgaaagtgtgacacacgtcttcaacatccgcaac | 4656 |

| | | | |
|--------|--|--|------|
| OY | 4277 | CTGACAGCGTTTCAGTCCGATVAGCAFTTMGGTGCTGCCAGCAGCATGTGTGAG | 4336 |
| | | | |
| Db | 4657 | ctcacaccatcatcgtactacacaagggtgcgtccttgagcaagaagaaatccaggag | 4716 |
| OY | 4337 | TTTGACACCCTGCCTCTCTCTG | 4360 |
| | | | |
| Db | 4717 | tacggcgcccatcgacctctg | 4740 |
| RESULT | 9 | | |
| ID | AAK19818 | | |
| | AAK19818 standard; cDNA: 5011 BP. | | |
| XX | AAK19818; | | |
| DT | 10-JUN-1999 | (first entry) | |
| DE | Human multidrug resistance-associated protein natural variant cDNA. | | |
| KW | Human; multidrug resistance-associated protein; MRP; cytotoxic drug; | | |
| KM | cancer; chemotherapy; ds. | | |
| OS | Homo sapiens. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | CDS | 196..4791 | |
| FT | | /tag= a | |
| PN | US5891724-A. | | |
| PD | 06-APR-1999. | | |
| XX | | | |
| PF | 05-JUN-1995; | 95US-0460907. | |
| XX | | | |
| PR | 05-JUN-1995; | 95US-0460907. | |
| PR | 27-OCT-1992; | 92US-0966923. | |
| PR | 08-MAR-1993; | 93US-0029340. | |
| PR | 26-OCT-1993; | 93US-0141893. | |
| PR | 20-MAR-1995; | 95US-0407207. | |
| PA | (TOOH) UNIV QUEBENS KINGSTON. | | |
| PI | Cole SPC, Deeley RG; | | |
| DR | WPI; 1999-253868/21. | | |
| DR | P-PsDB; AAW99894. | | |
| PT | Protecting mammalian cells against cytotoxic drugs | | |
| PS | Claim 4; Column 67-78; 82pp; English. | | |
| XX | | | |
| CC | The present sequence encodes a human multidrug resistance-associated | | |
| CC | protein (MRP). The present invention also describes a method for | | |
| CC | protecting a mammalian cell against the cytotoxicity of anthracyclines, | | |
| CC | epidodophyllotoxins and Vinca alkaloids (A) by introducing into it a | | |
| CC | nucleic acid (I) that hybridizes under stringent conditions to a nucleic | | |
| CC | acid (II) that encodes an MRP protein (III). Introduction of (I) | | |
| CC | normal cells against cytotoxic effects of (A), particularly to protect | | |
| CC | transformed with (I) can be used to screen for agents that affect | | |
| CC | multidrug resistance or are directly toxic to multidrug resistant cells, | | |
| CC | i.e. potential therapeutics for multidrug-resistant cancers. Conferring | | |
| CC | resistance to normal cells should allow an increase in the dose of (A) | | |
| CC | that can be administered safely. | | |
| XQ | Sequence 5011BP; 1064 A; 1498 C; 1394 G; 1055 T; 0 other; | | |

| | | | |
|----|------|---|------|
| QY | 1811 | CACATTCACCGGGCCACCCGCGCTTACACAGGACACCTGCACAGATCGATCGATGCAGATC | 1870 |
| Db | 2142 | cacatlcacccttggggcc--ggagcgaccctcccaaccctgaaatggaatcacctctccacc | 2139 |
| QY | 1871 | CMAAGGGGTAAACCTGGTGGGAAATCTCGGCGAGTGTGGGAATGGAAAAACCTCTTCAT | 1930 |
| Db | 2200 | ccggaagtgctcttggtagccgtgtagggccagatgtagctctggagaaatgctgcctgtgc | 2259 |
| QY | 1931 | TCAGCCATTTTAAAGCCAGATGACGCTTCTTAGAGGGCAGCATTTGCATACAGTGGAACTTC | 1990 |
| Db | 2260 | tcagccctcttgcttgatgtagcaagaatggaagggcagtgtagctatacaagggctcgtg | 2319 |
| QY | 1991 | GCTTATTTGGGCCACAGGCGCTGGATCTCCATGTCTACTCTAGAGACACAACTCCTGTT | 2050 |
| Db | 2320 | gctcatctgccaagcagggccttgatctcagaaatgatalctctccgagaataacatccctt | 2379 |
| QY | 2051 | GGGAAGGAATATGATGAAAGAAATACATCTGTGCTCMAACGCTGCTGCCTGAGGCTT | 2110 |
| Db | 2380 | ggatgctcagcttgtagggaaacatatataaggtccgtgatatcaagctctgycctctccca | 2439 |
| QY | 2111 | GACCTGGCCATTCTTCCACAGACGACCTGCAGGAGATTGGAGAGCGAGGACCAACTG | 2170 |
| Db | 2440 | gaccttggaataatccctgcccagtgggagatcgagacagatctggcagaaagggcgtgac | 2499 |
| QY | 2171 | AGCGGTGGGCAGCGCCACGAGGATCAGCCTTTGCCCGGCGCTTGTATATGATGACAGGAC | 2230 |
| Db | 2500 | ctctggggagacaagcagcgctgtagccctgcccggccgcttgatccaacgcttgacat | 2559 |
| QY | 2231 | TACATMCCGGACGACCCCTCAGTGCCTTAGATGCCCATGTGGGACACACATCTTCAT | 2290 |
| Db | 2560 | tacctctcgatctccctccctcctcagcagtgagatgaccaatgtaggaaacaacatcttgaa | 2619 |
| QY | 2291 | AGTGCAT-----CCGAAACATCTCAAGTCCACACAGATCTGTGTTTACCACACAG | 2344 |
| Db | 2620 | aatgtgattggcccccaagggatgctcgaagaacaagaacggtatcttgctcaagcagc | 2679 |
| QY | 2345 | TTACATGATACCTGGTTGACTGTGATGAAATGATCTTATATAAAGAGGCGTGTTCAGGA | 2404 |
| Db | 2680 | atgagcttactctgcgcgcagctgtagcgtcatcatctgatatgtagtgggcgcaagatctctga | 2739 |
| QY | 2405 | AGAGGACCCCTGAGGAACTGATGAATTTTAAATGGGACATATGCTACATCTTTTAAATAC | 2464 |
| Db | 2740 | atgggctccctaccgagagcgtcgtgctcgagaaagcgccctcgtcgaagtctcgtacc | 2799 |
| QY | 2465 | CTGTGCTGGAGAGACACCGCCAGTT--GAGATCAATTCAAAAAGGAAACCAAGTGT | 2521 |
| Db | 2800 | tatgcagcaaaagggcagggcaggaatcgaggaagaaacggggctcaagggcgctcaagcgct | 2859 |
| QY | 2522 | TCACAGAGGAAGTCAACAGACAG-----GTCTTAAACAGAGCATTAATAAAGG | 2573 |
| Db | 2860 | ccagggagaggaagaagaagaacaatggaagatgtagctgctgtagcaggaacagctgcaggaag | 2919 |
| QY | 2574 | AAAAAGCATTAACCCAGAGGAAGGCGACTTGTGACGTGG----- | 2615 |
| Db | 2920 | caactgcaagagacagctccacgacatcctcctcctataagtggggaacatcaagcaggaacac | 2979 |
| QY | 2616 | AAGAGAAAGGCGAGGTTCACTCCCTCGTGCATATATGTGTTCATATCCAGCTGCTG | 2675 |
| Db | 2980 | aacgcaacccgagagacctgcagaaagcttgaggccaagaabaagabagagacctggaaagctg | 3039 |
| QY | 2676 | GGGGC-----CCCTTGGCATCTCTGGTTATTATGAGCCCTTTTCAVCTGATGATGAGC | 2728 |
| Db | 3040 | gagctgtgacaagggcagacaagggcaggttaagcttctcgttactcgggaatclataagaa | 3099 |
| QY | 2729 | AGCAACCGCTTACACACTCGTGGTGTGAGTTACTGTGATCAACGAAGAAGCGGAAACAC | 2788 |
| Db | 3100 | gccaatcggaactctcatcctcctccatgaatcttcccttcaatgtaatgtaacaatgctgc | 3159 |
| QY | 2789 | ACTGTGACT-----CGAGGAACGAGACCTCGGTGAGTGAAGAGCATGAAGCAATCT | 2842 |
| Db | 3160 | gcgctgtgcttcaactatcttgctcagcctcttgacatgtagaccocatogtcaacggagct | 3219 |
| QY | 2843 | CATATGCACTACTATGCCAGCATCTACGCCCTTCCATGSGCAATCATGCTGATGCTGAAA | 2902 |


```

2561 tacctctcgtatgatccctctccagcagtgatgcccacgtgaggaaacacatctttgaa 2620
QY 2291 AGTGTAT-----CCGAAACATCTCAAGTCCAAAGACAGTTCTGTTTATACCCAG 2344
Db 2621 aatgfatctgcccacaagggatgctcgaagaacaagcgcgaatcttgctccgcagc 2680
QY 2345 TTACAGTACTGTTGACGTGATGATGAAGATCTTCAATGAAGAGGCGGTATACGGA 2404
Db 2681 atgagctactgcccaggtgagtcacatcgtcactgagtcgagcaagatcttgag 2740
QY 2405 AGAGCCACCCATGAGAGAACTGATGAATTTAAATGAGTATGCTACATTTTAAATAC 2464
Db 2741 atggctctctaccagagatgctgctcgagacggccttcgtgagtctctgctgac 2800
QY 2465 CTGTGCTGGGAGACACCCGCACTT---GAGATCAATTCAAAAAAGAAACCACTGAT 2521
Db 2801 tatgcacagcagagcagagatgacagatgacagggagaaacgggttcacgggcgtcagcggt 2860
QY 2522 TCACAGAAAGTACAAAGACAAG-----GTCTTAAACAGGATCAATTAAGAAAG 2573
Db 2861 ccagagaaagaaagcaaaagcaaatggaatgacatgctggtgacgagtcagggagag 2920
QY 2574 AAAAAGCATTAAGCCAGAGAGGACAGCTTGTCAGCTGG----- 2615
Db 2921 caactgcagagacagctcagcagctctctctctactatagtggggaactcagcagcaaac 2980
QY 2616 AAGAGAAAGGAGGAGGTTGATGCCCCGTGATGATGATGCTTCAATCCAGCTGCTG 2675
Db 2981 aacacacacgcagaaactgagaaactgagcacaagaaagagagacctgcagactgagctg 3040
QY 2676 GGGGC-----CCCTTGGCATTCCTGTTATGAGCCCTTTCATGCTGAATGAGGC 2728
Db 3041 gaggctgacaaggcagcagaaagcaggtcacaagctctccgtgctacgggaactacaagaa 3100
QY 2729 AGCACCGCTTCAGACACTGATGTTGATGATGATGATGATGATGATGATGATGATGATG 2788
Db 3101 gcatcgcagctctctatctctctctcagcatctctctctctctctctctctctctctctct 3160
QY 2789 ACTGTGACT-----CGAGGAAACGAGACCTGCGTGAAGCAAGCAATCCT 2842
Db 3161 ggcgtgactcccaactatgctcagctcctgactgactgactgactgactgactgactgactg 3220
QY 2843 CATATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2902
Db 3221 cagggacacacgaagaagtcggtcggtctatgagacccctgagcaatccacaagggatc 3280
QY 2903 GCCATTCGAG-----AGTTGTCTTTGTCAAGGCGACGCTGCGACCTTCTCCCGGCTG 2956
Db 3281 ggcgtgcttctgactcactcagtcgctgctcactcaggggagatctgctcccgctgctg 3340
QY 2957 CATGACGAGCTTTTCCGAAAGATCTCTGCAAGCCCTATTAAGTTTTCACAGACCCCC 3016
Db 3341 caegtggacctgctcacaagcactcgtgcagtcacacatgagctctctctctctctctctct 3400
QY 3017 ACAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3076
Db 3401 agtggaaacctggtgaaacctctctccagagagctggaacacgttgagccatcatgaccg 3460
QY 3077 TTTCAGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3136
Db 3461 gggatcatcaagatgtctcagtcgctccctctcctcctcctcctcctcctcctcctcctc 3520
QY 3137 GCAGAGAGTTCCTCCGAGTCTCTGTCGAGTGGGGCCCTGTGATCTCTTTTCAGTC 3196
Db 3521 ctctggtgcacagccatcgccgcatcactcactccgccccctctggtcctcactctctc 3580
QY 3197 CTGACATGCTCTCAGAGGCTCTGATTCGGAGCTGAAGCTGATGATGATGATGATGATGATG 3256
Db 3581 gtccagaggttcttaagtgtctctcccgagcagctggaagcgctcgtgacgtcagcgcg 3640
QY 3257 TCACCTTCTCTCCACATCAAGTCCAGCATTAAGGCGCTTGCACATTCACGCGCTAC 3316
Db 3641 tcccggtctattcccatctcaacgagacctgctggtgggtcagcgctcatctgagacctc 3700
QY 3317 AATTAAGGCGAGAGTTTTCACAGATACAGAGACTGCTGGATGATACAAAGCTCCT 3376
Db 3701 gaggagcagagacgctctcactccacagagtgacctgaaggtgagacgagacagaaagcgc 3760
QY 3377 TTTTTTTTCATGCTGTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3436
Db 3761 taattcccaagacatcgctgcccacaagtgagtcggccgtgagctgagatgctggtggaac 3820
QY 3437 GCCCTCATCACCACACAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3496
Db 3821 tgcctgctctgcttctgctcctcctgcttctgctgctgctcctcagcacaagctcagtgctgc 3880
QY 3497 TATGCGGCTCTGCCATCTTATGCTGTCAGTTTAAAGGGCTGTGTCATGTTTACGTC 3556
Db 3881 ttggtggtcctctcagtgcttactcactgacgtgacacacgtactgactgactgactgactg 3940
QY 3557 AGACGTGATGAGACAGAGCTGATTCACCTCGGTGAGAGATCAATCACTACAT 3616
Db 3941 cggatgctactcgaatggaatggaacacacatcgctgcccgtgagaggtcacaagatctca 4000
QY 3617 AAGACTGTCTCTTGGAGACACCTGCGAGATTAAGAACAGGCTCCCTCCCTGACTGG 3676
Db 4001 ggaactgagaa---ggaggcgccctgcaaatccagagagacagctccgcagcagctg 4057
QY 3677 CCCAGGAGGAGAGAGTGAATTTGAGAACGACAGAGATGATGATGATGATGATGATGATGATG 3736
Db 4058 ccccaagtggcgagtggaatctcgaactactgctgctgcactcgaagagagacgtgagc 4117
QY 3737 CTGCTCTTAAAGAGTATCTTCAAGTCAAACTTAAAGAGATGAGTGGGAG 3796
Db 4118 tctgtctcacaagcacaatcgaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaat 4177
QY 3797 CGAGACGATCAGGAGAGTCTCTGCGAGAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3856
Db 4178 cggagagagctgaggagagctgctccctgacccctggtctatcttgcagcaagctgctgc 4237
QY 3857 GAGGCTGATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3916
Db 4238 gaaagagagatcatcactgactgactgactgactgactgactgactgactgactgactgact 4297
QY 3917 AGCAAACTCTATCTATCTCTCAAGAGCGGCTGCTGATGAGTGAAGTGAAGTGAAGTGAAGT 3976
Db 4298 tccaagatcacatcatcctcccaagacccctgttctgttctggtctccctccogaagaaac 4357
QY 3977 TTGACCCCTTCAACAGTACACTGAAGACAGATTTGGAGTCCCTGAGAGAGACAC 4036
Db 4358 ctggaccatctcagccagactcagtgagaaagctggaagctggaagctggaagctggaagct 4417
QY 4037 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4096
Db 4418 ctgaaagactcgtgtcagccctctcgaagaagctagacacatgaatgacaaagcgag 4477
QY 4097 GATTACTTCTAGTGGGAGAGAGGAGCTCTGTCGATGATGATGATGATGATGATGATGATGATG 4156
Db 4478 gaggaaactcagtgctgagagcgccagctgtgctgctgctgctgctgctgctgctgctgag 4537
QY 4157 TGTAAATTCGATTTTAAAGAGGACAGCTGCCATGAGACAGACAGACTTATTTG 4216
Db 4538 acgaagactctgtgtgtgtagagcagcgccgtgagccgtgagaaacgagcagccctc 4597
QY 4217 ATTCAAGAACATCCGAGAAAGATTTGAGACTGATGATGATGATGATGATGATGATGATGAT 4276
Db 4598 atccagtcacacatccggacagagctcgaagactgacacgctcactcactcactcactc 4637
QY 4277 CTGACAGGTTCTAGGCTCCGATGAGATGATGATGATGATGATGATGATGATGATGATGATG 4336
Db 4658 ctcaacacacatcatgactgactcaacaaggtgtgactgctggaacaaggtgaaatccaggg 4717
QY 4337 TTTGACACCCCATTCGCTCTCTG 4360
Db 4718 taaggcgcccatcggaactcctg 4741
```



```

Db 3340 caagcggacatctgcaacagcatccgtcgatccacatagatctcttgagcggaccccc 3399
Oy 3017 ACAGGAGGATTTCTCAACAGGTTTCCAAAGACATGATGAAGTTGACTGGGCGTCCG 3076
Db 3400 agtggaaacctggtgaacacctctctccaaagagctggacaacatgacatccatccgc 3459
Oy 3077 TTCAGGCGGAGATGTTCTATCCAGAACGTTATCTGTTGTTCTGTTGTTGGAAATGATC 3136
Db 3460 gaggatcatcaagatgttcatctggtccctgttcaacgtcatctggtcgtcatctgtatc 3519
Oy 3137 GCAGGAGCTTCCGCTGAGTTCCTTGTGGCAGTGGGGCCCTTGTGATCTCTTTTCAGTC 3196
Db 3520 ctgcgtgccaagcccatcgccgcatcatcatccgcgcccttggtccatctactcttcc 3579
Oy 3197 CTGCACATTTGTTCTCAAGGTTCTGATTCGGGAGCTGAAGCTGTGACAAATATCAGCAG 3256
Db 3580 gtccaaaggtcttaactgcttctccctcgagctgaagcgtcgaagtcggtcgaagccgc 3639
Oy 3257 TCACCTTTCCTCCACATCAGTCCAGATACAGGGCTTGGCCACATCCAGCGCTTAC 3316
Db 3640 tccccggtcatcttccatcttcaacagacatctgtggtggtgagcgtcatctgaagcttc 3699
Oy 3317 AATTAAGGCGAGAGTTTCTGACAGATACAGAGCTGCTGATGACAAACAGCTCTCT 3376
Db 3700 gaggagcagagcgcttcatccacagagtgagctgaagtgagcagagacaagaagcgc 3759
Oy 3377 TTTTGTGTTTACGTGCGATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3436
Db 3760 taataccacagatcgttgcacacagtggtgcgtggtggtggtggtggtggtggtggtggt 3819
Oy 3437 GCCCTATTCACACACACGCGGCTGATGATGCTTCTTATCAGCGGAGATTCGCCAGCC 3496
Db 3820 tgcacgtctcgttctgcgcctgttgcgtgcatctccagcaacagctcagtgctgtgc 3879
Oy 3497 TATGCGGCTCTGCGCATCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3556
Db 3880 ttggtggtcctctcagtgcttcatcatctgacagtgatcaacagtgatctgacgtgctgt 3939
Oy 3557 AGACTGCGATCTGACAGACAGAGCTGATTCACCTCGGTGAGAGATCAATCACTACAT 3616
Db 3940 cggatgcatctgaatgaatgaacacacatcgtggtggtggtggtggtggtggtggtggt 3999
Oy 3617 AAGACTGCTCTTGGAGGACCTGCGACATTAAGAAAGAGCTCCCTGACTGCTG 3676
Db 4000 gaggactgagaa--ggaagcgccctggaacatccaggaacagctcgcgcgcgcgcgcgcgc 4056
Oy 3677 CCCAGAGAGGAGAGGTGACCTTTGAGAACGACAGATGAGGTACCGAAGAAACCTCCT 3736
Db 4057 ccccaagtggtgcgagtggaatctcgaactactgctcgtcgtacccgaagagacgtgac 4116
Oy 3737 CTGCTCTTAAGAAAGTATCTTCAAGATCAAACTAAAGAGAGATTGGCTGTGGG 3796
Db 4117 ttctgtctcaagacatcatgcatgcaatcgaatcgaatcgaatcgaatcgaatcgaatcga 4176
Oy 3797 CGGACGAGATCAGGAGAGCTCGCTGGGGATGGCCCTCTCGCTGCTGCTGCTGCTGCTGCT 3856
Db 4177 cggatcggaagctcggaagagctccctcgaccccggtcttctcgaatcaacagatctgc 4236
Oy 3857 GGAGGCTGATCAAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3916
Db 4237 gaaagagagatcatcatctgcatcaacatcgcgaagatcgcgtcgtcgaagacccgcgc 4296
Oy 3917 AGCAAACTCTATATCTATCTCTCAAGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3976
Db 4297 ttcaagatcatcatctccatcccaagacccgttctgttctcgaatcgaatcgaatcgaat 4356
Oy 3977 TTGACACCCCTTCAACAGATCACTGAAGACAGATGATGATGATGATGATGATGATGATGAT 4036
Db 4357 ctggaacatctcaagcagctacccgaatgaagaagctcgtgagctcgtgagccccc 4416
Oy 4037 ATGAAAGAAATGATTGCTTCACGCTACCTCTGAAGCTTGAATCTGAAGTGAATGAGG 4096

```

```

Db 4417 ctgaagactctgtcagccctctcgaacagctagacatgaatgtgcagaagcgagg 4476
Oy 4097 GATPACTTCTGATGGGGAGAGGAGCTCTTGTGATGATGATGATGATGATGATGATGATGAT 4156
Db 4477 gagaacctgaatgtcggagcagccagctctgtcgtcagcccggtcgtcgtcgtcgtcgtc 4536
Oy 4157 TGTAAATTTCTATTTTATATGAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4216
Db 4537 acgaagatctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4596
Oy 4217 ATTCAGAGACCATCCGAGACAGATTTGCAAGCTGATCCATGCTGCAATGCTGCTGCTGCT 4276
Db 4597 atccatccacatcccgacacagttcgaggaatgacccgtctcaccatccgcgcgcgcgc 4656
Oy 4277 CTGCACACGCTTCTAGGCTCCCATGATGATGATGATGATGATGATGATGATGATGATGAT 4336
Db 4657 ctcaacacatcatgtgacacacagtgatcgtctcttgacaagaagaaatccagag 4716
Oy 4337 TTTGACACCCCATCGGCTCCTG 4360
Db 4717 taagcgcgccatcggaactctgt 4740

```

```

RESULT 13
AA239556
ID AA239556 standard; cDNA; 5011 BP.
XX
AC AA239556;
XX
DT 28-FEB-2000 (first entry)
XX
DE Human MRP variant encoding cDNA.
XX
KW Chemosensitizer; multidrug resistance-associated protein; MRP; human;
KW therapeutic agent; P-glycoprotein-mediated multidrug resistance; lung;
KW cancer; variant; ds.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 196..4791
FT mutation 2249
FT /*tag= a
FT /*product= "MRP variant"
FT mutation 4039
FT /*tag= b
FT /*note= "wild-type T is changed to C"
FT mutation 4040
FT /*tag= c
FT /*note= "wild-type C is changed to G"
FT mutation 4040
FT /*tag= d
FT /*note= "wild-type G is changed to C"
XX
US6001563-A.
PD 14-DEC-1999.
XX
PF 05-JUN-1995; 95US-0463179.
XX
XX 27-OCT-1992; 92US-0966923.
XX 08-MAR-1993; 93US-0029340.
XX 26-OCT-1993; 93US-0141893.
XX 20-MAR-1995; 95US-0407207.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Cole SP, Deeley RG;
XX
DR WPT: 2000-061877/05.
XX P-PSDB: AA155799.
XX
Identification of chemosensitizers useful for treating cancer, using

```



```

Db 4000 gagactgagaa--gagagcgccctggaatccagagacagctccgccagcagctg 4056
QY 3677 CCCCGAGGAGGAGAGTGAACCTTTGAGAACGACAGATAGAGTACGAGAAACCTCCCT 3736
FT ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
FT 4057 ccccgagtgagcgagtggaatccagaaactactgctgcgtagcagagagagcctgagc 4116
FT 3737 CTCGTCCTTAAGAAAGTATCTTCACGATCAAAACCTAAGAGAGATGCGCTTGGGG 3796
Db 4117 ttcgtctcaggaacatcaatgcatcagatcaatgaggaagaaagtcgagcagctg 4176
QY 3797 CGGACGAGATCAGGAGAGTCTCGCTGGGATGCGCCCTTCCTCGTGGTGAATATCT 3856
Db 4177 cggacggagcgaggaagtcgctccctgaccctggtcttatccgagcaagctgac 4236
QY 3857 GGAGGCTGATCAAGATTCATGAGTGAATCAGTGAATTCGCTCCGACCTCCCA 3916
Db 4237 gaagagagatcatcatgcatcgaacatcgcaaatcgagctgacgacctccgc 4296
QY 3917 ACCAACTCTATCATCTTCCTCAAGAGCGGTCGTGTGAGTGACACTGATCAAT 3976
Db 4297 tccaagatcaccatcatccccaagacccctgttctgttctgggtccctccgaatgaac 4356
QY 3977 TTGACCCCTTCAACCATGACTGTAAGACAGATTTGGGATGCCCTGAGAGACACAC 4036
Db 4357 ctggaaccatcagcagactcagatgagaagctgagcgtcgtgagctgagctccac 4416
QY 4037 ATGAAAGATGATTCGCTACGTACTCTGAACCTTGAACTGGAATGATGAGATGGG 4096
Db 4417 ctgaagagactgctgacgacctctctgacaagctagaccatgaatgtagaagcgag 4476
QY 4097 GATTAATCTCTCACTGGGAGACGCACTCTTGTGATGACTAGAGACCCCTGCCGCCAC 4156
Db 4477 ggaacacctagctgctgagcgagcgccagctgtgctcctgagcccgccctgtgagag 4536
QY 4157 TGTAAATCTGATTTAGATGAGACGACAGCTGCCATGAGACAGACAGACTTATTG 4216
Db 4537 acgaagatcctgtgtgtgtagagcagcgagccgctgacctgtaaacgagccac 4596
QY 4217 ATTCAAGAGACCATCGAGAACATTTGACACTGTACACTGCTACCATTCGCATGCC 4276
Db 4597 atccagatccacatccgagacagactcgagactcaccgctccacatcgccacgcg 4656
QY 4277 CTGACACGCTTCTAGGCTCCGATGAGATTAAGTGTGCTGCCACGAGAGGAGGAG 4336
Db 4657 ctcaacacatcatgtagacacaaagtgatcgtcttgagaaagagaaacacagag 4716
QY 4337 TTTGACACCCCATCGCTCTTCTG 4360
Db 4717 tacggcgcccatcgacctcctg 4740

```

```

RESULT 14
AAFI4911
ID AAFI4911 standard; cDNA; 4864 BP.
AC AAFI4911;
XX 08-AUG-1996 (first entry)
DE cDNA encoding multi-drug resistance protein alternative splice form 2.
XX
XX MRP, multi-drug resistance protein; small cell lung cancer; protect.
XX cytotoxicity; therapy; P-glycoprotein; alternative splice variant; ss.
OS Homo sapiens.
FH Key 1.195 Location/Qualifiers
FT 5'UTR 1.195
FT CDS 196..4644
FT /*tag= b
FT /*product= multi-drug_resistance_protein
FT /*note= "alternative splice form"

```

```

FT misc_difference 1845..1846
FT /*tag= c
FT /*note= "nucleotides 1846-1992 of AAFI1713 deleted to
FT maintain ORF, not 1845-1992 as indicated in the
FT specification"
FT 3'UTR 4644..4864
FT /*tag= d
XX US5489519-A.
XX
XX 06-FEB-1996.
XX
XX 27-OCT-1992; 92US-0966923.
XX
XX 26-OCT-1993; 93US-0141893.
XX 27-OCT-1992; 92US-0966923.
XX 08-MAR-1993; 93US-0029340.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
XX
XX Cole SPC, Deeley RG;
XX
XX WPI; 1996-115615/12.
XX P-PSDB; AAR96953.
XX
XX Multi-drug resistance protein, MRP and DNA encoding it - for
XX diagnosis and treatment of e.g. small lung cell cancers
XX
XX Claim 27; Column -: 50pp; English.
XX
XX The present sequence is an alternative splice form of a gene (AAFI1713)
XX which encodes a multi-drug resistance protein (MRP). The specification
XX states that nucleotides 1845-1992 are removed (amino acids 551-559
XX inclusive of AAR9153), however this does not maintain the reading
XX frame, hence the sequence presented here has nucleotides 1846-1992
XX removed. The gene was isolated from H69AR, a small cell lung cancer cell
XX line, which is about 50-fold resistant to doxorubicin as compared to the
XX parental H69 cell line. The human MRP gene has been mapped to 16p13.1,
XX and is not linked to either CTR or MDR genes. The MRP protein is
XX overexpressed in multi-drug resistant cells independently of
XX overexpression of P-glycoprotein (a member of a superfamily of membrane
XX proteins that serve to transport a variety of mols., ranging from ions to
XX proteins, across cell membranes, and associated with a variety of
XX tumors). The nucleic acids can be used to protect a drug sensitive cell
XX from cytotoxicity due to exposure to a drug, or to develop improved forms
XX of therapy, particularly for small cell lung cancers which exhibit a high
XX initial response to chemotherapy which is almost always followed by a
XX multi-drug resistance form of the disease.
XX
XX Sequence 4864 BP; 1040 A; 1449 C; 1363 G; 1012 T; 0 other:

```

```

Query Match 10.4%; Score 506.4; DB 17; Length 4864;
Best Local Similarity 52.4%; Pred. No. 3.5e-120;
Matches 1364; Conservative 0; Mismatches 1181; Indels 59; Gaps 9;
QY 1811 CACATCCACCTGGGCCACCTGCGCTTACAGAGACACTGCACAGCATGATCGAGATC 1870
FT ||||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
FT 1995 cacattcacctggcca--ggaagcgacccctccacacactgaatgacatccatccatc 2052
QY 1871 CAAGAGGCTAACTGGTGTGGAATCTGCGCAGTGTGGAAGGAAAGTGAACCTCTCAT 1930
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 2053 ccgaagagtgcttgctgctgctgctgctgctgctgctgctgctgctgctgctgctc 2112
QY 1931 TCAGCCATTTTAAAGCCAGATGACGCTTCTAGAGAGGCGACGATTCATGATGAACTTC 1990
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 2113 tgaagcctcttgctgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 2172
QY 1991 GCTTATGTGGCCAGCAGGCTGTGATCTCAATGCTACTCTGAGAGACAACTCTGTTT 2050
FT ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 2173 gctatgtgccacagcagcgctgagatgactctcccgtaaaacatccctttt 2232
QY 2051 GGAAGGATATGATGAGAAAGATCAACTCTGTGCTGAACAGCTGCTGAGAGGCT 2110

```


| | | | |
|----|------|---|-------|
| D | 2794 | caactgagagagacagctcagcagctctctctctactatagtaggggacatcagcaagcaac | 28533 |
| OY | 2616 | AAGAGAAAGGCGAGGGTTACAGTCCCTGGTCAGATATAGGTCTTACATCCAGGCTGCTG | 26755 |
| D | 2854 | aacagacacccgaactcgcgaagaagctctgagcccaagaagagagactctggaactgtag | 29133 |
| OY | 2676 | GGGCG-----CCCTGGCATCTCCGGTTATTTATAGGCCCTTTTCATCTGATAGTAGGC | 27288 |
| D | 2914 | gagcgctgacaaggcgcaagagagcgaagctcgaagcttccggtactggtactacatgaag | 29733 |
| OY | 2729 | AGCACCGCCCTTCAGCACACCTGCTGGTTGAGTTACTGTGATCAGAACGAAGAGCGGGAACCC | 27888 |
| D | 2974 | gcactcgactcttcaatctctctctccagcatctctcttcaigtatgaacatgtgc | 30333 |
| OY | 2789 | ACTGTACT-----CGAGGGAACGAGACCTCGGTGAGTGACATGACATAGGACATCT | 28422 |
| D | 3034 | gcgcctgcttccaactatgtgctcgaactctgtgactgtagtaccatcgtaacacggagct | 30933 |
| OY | 2843 | CATATGACGATCATATGCGACGATCTACGGCCCTCTCCATGCGCATGCTGATCTCTGAA | 29022 |
| D | 3094 | cagagagacacgaagatccggcctgagcgtctatgagccctggagctttcaacaaggatc | 31533 |
| OY | 2903 | GCCATTGAGAG-----AGTTGCTTTGTTCAAAGGGGACGCTCGAGCTCTCCGGCTG | 29566 |
| D | 3154 | gcgcgtgtgtgctaccccaatgcccgtgtccatccggggatcttgcttccgcctgtctg | 32133 |
| OY | 2957 | CATGACGAGCTTTTCCGAGAGATCTTCGAAAGCCTATGAGATTTTGTGACACGACCCC | 30166 |
| D | 3214 | caagtgaactcgtgcacagcatctcgtgcgtacacatagactcttttgaagggccccc | 32733 |
| OY | 3017 | ACAGGGAGATTTCTCAACAGGTTTTCAAAGACATNGAGATGAGTGTAGCTGGCGTGGCG | 30766 |
| D | 3274 | agtggagaaacttggacacgcgtcttccaagagagctggaacacagtgactcaatgctccg | 33333 |
| OY | 3077 | TTCCAGGCGCCAGATGTTCACTCAGAACTGTATCCGTGCTGTTCTGTGTGGGATGAT | 31366 |
| D | 3334 | gaggtatcaagatgttcaatgagctctccctgtccaagatcatgtgctcgtcatgctatc | 33933 |
| OY | 3137 | GCAAGAGCTCTTCCCGTGGTTCTTGTGGCAGTGGGGCCCTTGTATCTCTTTTCAGTC | 31966 |
| D | 3394 | ctgtgtgcacagcccaatcgcgcgacatcatcatccgcgcctgtgctcatctactcttc | 34533 |
| OY | 3197 | CTGACACTTGTCTCCACAGGGGCTCGATTGGGGAGCGTAAGCGCTGCGACAAATACCGAG | 32566 |
| D | 3454 | gtccagaagttctcaatggtcttccctccgcgcagctgaaagcgcttcgaatcgtgcacgc | 35133 |
| OY | 3257 | TCACCTTTCCCTCCACATCATGACGTCCACATACAGGCGCTTGCACACATCCAGCTTAC | 33166 |
| D | 3514 | ttcccggtcatcttccatcttcaacgagactcgtcgggtgaagtgatctcagcctc | 35733 |
| OY | 3317 | AATTAAGGCGAGAGTTTCTGCACAGATATCCAGAGAGCTGCGATGACAAACAGTCTCT | 33766 |
| D | 3574 | gagagagcagagcgtcttcaatccacagagtgacatcgaagtgtagcagaaacgaagc | 36333 |
| OY | 3377 | TTTTTTTTTTTAAAGTGTGGAGTGGCGGTGGCGGTGCTGCGGCTGACCTCATACGATC | 34366 |
| D | 3634 | tattaccaccagatcgtlbgcacaagtggtgcgtggtgagtggtgtggtgagac | 36933 |
| OY | 3437 | GCCCTCATCACACCAAGGGGCTGATGATCGTCTTATTCACAGGGCAATTTCCCAACCC | 34966 |
| D | 3694 | tgcactgtctgttgctgctgcctcgttgcgtggtatccagacaaagctcaggtgtcgc | 37533 |
| OY | 3497 | TATGCGGCTTCGCCATCTCTTAAGTGTGTCAGTTAACGGGGCTTTCAGATTACGGCTC | 35566 |
| D | 3754 | ttgtgtggtcctcagtgcttactcatctatgacagtgtaacacgttacttgactggtgtt | 38133 |
| OY | 3557 | AGACTGGCATCTGAGACAGAAGCTCGATTCACCTCGGTGGAGAGATCAATCACTACTT | 36166 |
| D | 3814 | cggatgtcatctgtaaatgaaacaaacatcgtgcgtgtgagaggtctcaagagatattca | 38733 |
| OY | 3617 | AAGACTGTGCTCTGGAAGACCTGCGCAAAATTAAGAACAAAGGCTTCCCTCCATGACGG | 36766 |
| D | 3874 | gagagctgagaa---ggagagggccttgcgaataaccagaaacacgtctccgcacgaagctgg | 39320 |

| | | | |
|----|------|---|-------|
| QY | 3677 | CCCCAGAGGGAGGACCTTTGGAAACGAGATAGAGGTACCGAAACCTCCCT | 37368 |
| Db | 3991 | ccccagatggcccgagatggaattccggaatactactctccgctcgcatacgaagaagacttgac | 3990 |
| QY | 3737 | CTCCTCTTAAAGAAAGTATCTTTCACGATCAAACTTAAAGAGAAGATTGGCATTTGGGG | 37966 |
| Db | 3991 | ctcgtcttcagcgcatacctaattgcacgatctaattgagggaagaagctgcgacatcgtaggg | 4050 |
| QY | 3797 | CGGACAGATCAGGGAAGTCTCGCGGGGATGGGCCCTTCCTCGCTGSGTAGTTACT | 38566 |
| Db | 4051 | cggacggagactgggaagctgtctcccggaaccctgggacttaattcggatcaacagatctgc | 4110 |
| QY | 3857 | GGAGGCTGCATCAAGATTGGATGGAGTGAAGTATAGTATTTGGCTTCCGACCTCCGA | 3916 |
| Db | 4111 | gaaggagagatcatcatcgtatgctgcatacaatcgcgaagatcgcgctgaagacctccgc | 4170 |
| QY | 3917 | AGCAAACTCTCTATCATCTCTCAAGACCGGCTGCTGTATAGTGGCATTGTAGATCAAT | 3976 |
| Db | 4171 | ttcaagatccaccatcatcccccagagaccctgtttgttttcgggttccctccgaatgac | 4230 |
| QY | 3977 | TTGGACCCCTTCAACCGATACACTGAAAGACATTTGGGATGCCCTGGAGGACAC | 40366 |
| Db | 4231 | ctggaaccttccagccagctactcctgagtgaagaagcttggagctccctgagctgcccac | 4230 |
| QY | 4037 | ATGAAAGATGTATTGCTCAGCTACTCTGAAACTTGAACTTGAAATGTGAATGAGATGCG | 40966 |
| Db | 4291 | ctgaagagactctgctgaagccctctcctgaacagctagacaaatgaatgtcagaagcggg | 4350 |
| QY | 4097 | GATPACCTTTCAGTGGGGGGGAAAGGACGCTTGTGATAGTGAAGACCTGCTCCGCCAC | 41566 |
| Db | 4351 | gagaaacctccagctgcgggacgacgacgcttgtgtgctcgaagcccgagccctgcctgaggaag | 4410 |
| QY | 4157 | TGTAAAGATTCTGATTTTATGATGAAGACACAGCTCCATGAGACAGACACTTATTG | 4216 |
| Db | 4411 | acgaaagatccttctgtgtgtgtaggacagcgacgacgctgtgaacctgaaacgaacgacctc | 4470 |
| QY | 4217 | ATTGACAGACATCCGAGAAAGATTGGACATGTGACATCTGTACCATTTGCCATGCG | 4276 |
| Db | 4471 | atccagctacacacacgagacacagcttcgaagatgcacacgcttcctacatctgcgccacg | 4530 |
| QY | 4277 | CTGCACACAGGTTTCAAGCTCCGATAGGATTTATGTCCTGCCACAGGACAGGTTGTTGAG | 4336 |
| Db | 4531 | ctcaacacatcatctgagctacacaagaagtgatcgtcttgcacaagaaggaatccagag | 4590 |
| QY | 4337 | TTTGACACCCCATCGGCTCTTTCG | 4360 |
| Db | 4591 | tacggcgcccatatcgagctccctcg | 4614 |

Search completed: November 9, 2001, 15:58:48
Job time: 10446 sec

Oy 3617 AAGACTCTGTCTTGGAGCACCTGCGCAGAATTAAACAAGCGCTCCCTGCCCTGACTGG 3676
||||| - ||||| - ||||| - ||||| - |||||
Db 3874 ggagactggaaa---ggagcgccctgcgaatccagggagacacgttcgccccaagcagctcgg 3930

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2001, 12:19:42 ; Search time 153.99 Seconds
(without alignments)
7128.637 Million cell updates/sec

Title: US-09-528-031-1

Perfect score: 4847
Sequence: 1 GGCTCATGCTCGGAGCGGTG.....AAAAAAAAAGCGCGCCG 4847

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCrUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|------------------|
| 1 | 4847 | 100.0 | 4847 | 3 | US-09-061-400-1 |
| 2 | 4775.4 | 98.5 | 4781 | 2 | US-09-001-273-1 |
| 3 | 4775.4 | 98.5 | 4781 | 4 | US-08-843-459A-1 |
| 4 | 511.2 | 10.5 | 5011 | 1 | US-08-463-092B-3 |
| 5 | 511.2 | 10.5 | 5011 | 2 | US-08-462-109A-3 |
| 6 | 511.2 | 10.5 | 5011 | 2 | US-08-460-907B-3 |
| 7 | 511.2 | 10.5 | 5011 | 3 | US-08-463-179A-3 |
| 8 | 511.2 | 10.5 | 5011 | 3 | US-08-463-384B-3 |
| 9 | 506.4 | 10.4 | 5011 | 1 | US-08-141-893-1 |
| 10 | 506.4 | 10.4 | 5011 | 1 | US-08-463-092B-1 |
| 11 | 506.4 | 10.4 | 5011 | 2 | US-08-462-109A-1 |
| 12 | 506.4 | 10.4 | 5011 | 2 | US-08-460-907B-1 |
| 13 | 506.4 | 10.4 | 5011 | 3 | US-08-463-179A-1 |
| 14 | 506.4 | 10.4 | 5011 | 3 | US-08-461-384B-1 |
| 15 | 506.4 | 10.4 | 5011 | 3 | US-08-407-207A-1 |
| 16 | 420.4 | 8.7 | 5889 | 1 | US-08-463-092B-5 |
| 17 | 420.4 | 8.7 | 5889 | 2 | US-08-462-109A-5 |
| 18 | 420.4 | 8.7 | 5889 | 2 | US-08-460-907B-5 |
| 19 | 420.4 | 8.7 | 5889 | 3 | US-08-463-179A-5 |
| 20 | 420.4 | 8.7 | 5889 | 3 | US-08-461-384B-5 |
| 21 | 370 | 7.6 | 4931 | 4 | US-08-726-320-2 |
| 22 | 370 | 7.6 | 4931 | 4 | US-09-208-716-2 |
| 23 | 346 | 7.1 | 4877 | 2 | US-08-404-531B-7 |
| 24 | 346 | 7.1 | 4877 | 2 | US-08-404-531B-8 |
| 25 | 346 | 7.1 | 4877 | 3 | US-08-476-900A-7 |
| 26 | 346 | 7.1 | 4877 | 3 | US-08-476-900A-8 |
| 27 | 346 | 7.1 | 4877 | 3 | US-08-488-546A-7 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 346 | 7.1 | 4877 | 3 | US-08-488-546A-8 | Sequence 8, Appl1 |
| 29 | 344.4 | 7.1 | 5175 | 4 | US-08-972-927-4 | Sequence 4, Appl1 |
| 30 | 342.6 | 7.1 | 5110 | 2 | US-08-404-531B-4 | Sequence 4, Appl1 |
| 31 | 342.6 | 7.1 | 5110 | 2 | US-08-404-531B-5 | Sequence 5, Appl1 |
| 32 | 342.6 | 7.1 | 5110 | 2 | US-08-476-900A-4 | Sequence 4, Appl1 |
| 33 | 342.6 | 7.1 | 5110 | 3 | US-08-476-900A-5 | Sequence 5, Appl1 |
| 34 | 342.6 | 7.1 | 5110 | 3 | US-08-488-546A-4 | Sequence 4, Appl1 |
| 35 | 342.6 | 7.1 | 5110 | 3 | US-08-488-546A-5 | Sequence 5, Appl1 |
| 36 | 340.8 | 7.0 | 463 | 2 | US-09-001-273-3 | Sequence 3, Appl1 |
| 37 | 340.8 | 7.0 | 463 | 2 | US-09-061-400-3 | Sequence 3, Appl1 |
| 38 | 340.8 | 7.0 | 463 | 4 | US-08-843-459A-3 | Sequence 3, Appl1 |
| 39 | 336.2 | 6.9 | 5232 | 4 | US-08-972-927-1 | Sequence 1, Appl1 |
| 40 | 330.8 | 6.8 | 2454 | 2 | US-08-404-531B-32 | Sequence 32, Appl1 |
| 41 | 330.8 | 6.8 | 2454 | 3 | US-08-476-900A-32 | Sequence 32, Appl1 |
| 42 | 330.8 | 6.8 | 2454 | 3 | US-08-488-546A-32 | Sequence 32, Appl1 |
| 43 | 294 | 6.1 | 2294 | 2 | US-08-404-531B-30 | Sequence 30, Appl1 |
| 44 | 294 | 6.1 | 2294 | 3 | US-08-476-900A-30 | Sequence 30, Appl1 |
| 45 | 294 | 6.1 | 2294 | 3 | US-08-488-546A-30 | Sequence 30, Appl1 |

ALIGNMENTS

```
RESULT 1
US-09-061-400-1
: Sequence 1, Application US/09061400
: Patent No. 6077936
:
: GENERAL INFORMATION:
: APPLICANT: SHYUAN, Andrew
: TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE ASSOCIATED
: POLYPEPTIDE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/061,400
: FILING DATE: 16-APRIL-1998
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth A. Hanley
: REGISTRATION NUMBER: 33,505
: REFERENCE/DOCKET NUMBER: MNT-056CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 742-4214
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4847 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 116..4426
:
: US-09-061-400-1
:
: Query Match 100.0%; Score 4847; DB 3; Length 4847;
: Best Local Similarity 100.0%; Pred. No. 0;
: Matches 4847; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: 1 GGCTCATGCTCGGAGCGGTGAGCGGCTGCGGCTGCTGAGACGAGCGGCGAG 60
: |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

| | | | | |
|---|---|------|---|------|
| D | b | 1 | GGCTCATGCTCGGAGAGGTGGTTGAGCCGGCTGCGCGGGTGGCTTGAGAGCAGGGGAGCGAG | 60 |
| Q | y | 61 | GAATTCGATGTGAACACTAACAAGTCTGTGATAGCCCTTGGAAACCTCCACTAGAGAAAGATGAA | 120 |
| D | b | 61 | GAATTCGATGTGAACACTAACAAGTCTGTGATAGCCCTTGGAAACCTCCACTAGAGAAAGATGAA | 120 |
| Q | y | 121 | GGATATGACATAGAAAAAGATATATCATCTCCAGTCCCTGGGTGTAAAGATGTGAGGGA | 180 |
| D | b | 121 | GGATATGACATAGAAAAAGATATATCATCTCCAGTCCCTGGGTGTAAAGATGTGAGGGA | 180 |
| Q | y | 181 | GAGAACACAGACATTTCTGGGACGCACAGAGAAGCTGATGATTCAAAGTTCTAGAGAAACTCG | 240 |
| D | b | 181 | GAGAACACAGACATTTCTGGGACGCACAGAGAAGCTGATGATTCAAAGTTCTAGAGAAACTCG | 240 |
| Q | y | 241 | ACCGTTGGATGCCAAGATACCTTTGGAAACAGACAGCCGAGCCGAGGSCCTCTCTCTTGA | 300 |
| D | b | 241 | ACCGTTGGATGCCAAGATACCTTTGGAAACAGACAGCCGAGCCGAGGSCCTCTCTCTTGA | 300 |
| Q | y | 301 | TGCGCTCATGCAATTCACAGCTCAGAAATCCTGGATGAGAGACATCCCAAGGAAAGTACA | 360 |
| D | b | 301 | TGCGCTCATGCAATTCACAGCTCAGAAATCCTGGATGAGAGACATCCCAAGGAAAGTACA | 360 |
| Q | y | 361 | TGATAGGCTTAGAGCTCTGAGAGCCATCCGGACACTCTTGCAACACAGACCCCAAGTGA | 420 |
| D | b | 361 | TGATAGGCTTAGAGCTCTGAGAGCCATCCGGACACTCTTGCAACACAGACCCCAAGTGA | 420 |
| Q | y | 421 | CAATGCTGGGCTTTTTTCCTGATATGACTTTTTTCGTGGCTTTCTCTGAGCCGCTGTGGC | 480 |
| D | b | 421 | CAATGCTGGGCTTTTTTCCTGATATGACTTTTTTCGTGGCTTTCTCTGAGCCGCTGTGGC | 480 |
| Q | y | 481 | CCAAACAAGAGGGAGAGCTCTCAATGGAACAGCTGTGGTCTCTGTCCAGACAGAGTCTC | 540 |
| D | b | 481 | CCAAACAAGAGGGAGAGCTCTCAATGGAACAGCTGTGGTCTCTGTCCAGACAGAGTCTC | 540 |
| Q | y | 541 | TGACGTAAGTGCAGAGAAGACTATAGAGACTGTGGCAAGAGAAGCTGAATGAATGGGGCC | 600 |
| D | b | 541 | TGACGTAAGTGCAGAGAAGACTATAGAGACTGTGGCAAGAGAAGCTGAATGAATGGGGCC | 600 |
| Q | y | 601 | AGAGCTGTCTCCCTGGGAAGGTTGTGTGATCTTCTCTGCGCACCAAGGCTCATCTGTCTC | 660 |
| D | b | 601 | AGAGCTGTCTCCCTGGGAAGGTTGTGTGATCTTCTCTGCGCACCAAGGCTCATCTGTCTC | 660 |
| Q | y | 661 | CATGTTGCTGCTGATGATACAGAGAGTGGCTGAGCTCAATGAGACAGGCTTCATGTGAA | 720 |
| D | b | 661 | CATGTTGCTGCTGATGATACAGAGAGTGGCTGAGCTCAATGAGACAGGCTTCATGTGAA | 720 |
| Q | y | 721 | ACACCTTTGGAATATACCCAGGCAACAGAGTCTAACTGTGACATACAGCTTTGTTAGT | 780 |
| D | b | 721 | ACACCTTTGGAATATACCCAGGCAACAGAGTCTAACTGTGACATACAGCTTTGTTAGT | 780 |
| Q | y | 781 | GCTGGGCTCTCTCTGACGGAATCGTGGGGCTTGGTGCATGACATGAGTGGGAT | 840 |
| D | b | 781 | GCTGGGCTCTCTCTGACGGAATCGTGGGGCTTGGTGCATGACATGAGTGGGAT | 840 |
| Q | y | 841 | GAATTACCGAACCAGGTGTCGGCTTGGGGGGGGCCATCCTAACATAGGATTTAAGAAT | 900 |
| D | b | 841 | GAATTACCGAACCAGGTGTCGGCTTGGGGGGGGCCATCCTAACATAGGATTTAAGAAT | 900 |
| Q | y | 901 | CCTTAAGTTAAAGACATTAAAGAAATCCTGGGTGAGCTCATCAACATTTGCTCAA | 960 |
| D | b | 901 | CCTTAAGTTAAAGACATTAAAGAAATCCTGGGTGAGCTCATCAACATTTGCTCAA | 960 |
| Q | y | 961 | CGATGGGCAAGAAATGTTTGAAGGCAACACCGTGGAGCCGTGGTGGTGAAGAACCGT | 1020 |
| D | b | 961 | CGATGGGCAAGAAATGTTTGAAGGCAACACCGTGGAGCCGTGGTGGTGAAGAACCGT | 1020 |
| Q | y | 1021 | TGTTGCATCTTAGGATGATTTATTAATTAATTAATTCGTGGAGCCAAAGGCTTCTGGG | 1080 |
| D | b | 1021 | TGTTGCATCTTAGGATGATTTATTAATTAATTAATTAATTCGTGGAGCCAAAGGCTTCTGGG | 1080 |
| Q | y | 1081 | ATCAGCTGTTTTATCTCTTTTACCAGCAATGATGTTTGATACAGGCTCACAGCAT | 1140 |
| D | b | 1081 | ATCAGCTGTTTTATCTCTTTTACCAGCAATGATGTTTGATACAGGCTCACAGCAT | 1140 |

| | | | |
|----|------|--|------|
| QY | 1141 | TTTACGGAGAAAATGCGTGGCCGCCAGGATGAACTGTCCAGAAAGATGAATGAAGTCT | 1200 |
| Db | 1141 | TTTCCGGAGAAAATGCGTGGCCGCCAGGATGAAGTGTCCAGAAAGATGAATGAAGTCT | 1200 |
| QY | 1201 | TACTTACATTTAAATTTATCAAAATGTATGCTGGGTCAAGACATTTTCTCAGAGTGTCA | 1260 |
| Db | 1201 | TACTTACATTTAAATTTATCAAAATGTATGCTGGGTCAAGACATTTTCTCAGAGTGTCA | 1260 |
| QY | 1261 | GAAATTCGGCAGAGGAGGAGCGCTGGATTTGTGAAAACCCGGGTACTCTCCAGAGCATAC | 1320 |
| Db | 1261 | GAAATTCGGCAGAGGAGGAGCGCTGGATTTGTGAAAACCCGGGTACTCTCCAGAGCATAC | 1320 |
| QY | 1321 | TGTGGGTGTGGCTCCCATTTGTGTTGTTATTTGCCAGCTGGTGACCTTCTGTTCATAT | 1380 |
| Db | 1321 | TGTGGGTGTGGCTCCCATTTGTGTTGTTATTTGCCAGCGTGGTGACCTTCTGTTCATAT | 1380 |
| QY | 1381 | GACCTTGGGCTTCGATCTGACAGCAGCAGGCTTTCACATGGTGGACAGTCTTCATTC | 1440 |
| Db | 1381 | GACCTTGGGCTTCGATCTGACAGCAGCAGGCTTTCACAGTGGTGGACAGTCTTCATTC | 1440 |
| QY | 1441 | CATGACTTTTGTGTTGAAAGTAAACCCGTTTTCAGTAAAGTCCCTCAGAAAGCTTCAGT | 1500 |
| Db | 1441 | CATGACTTTTGTGTTGAAAGTAAACCCGTTTTCAGTAAAGTCCCTCAGAAAGCTTCAGT | 1500 |
| QY | 1501 | GCGTGTTCACAGATTTAAAGCTTGTGTTTCTAATGAGAGAGGTTCCATGATTAAGACAA | 1560 |
| Db | 1501 | GCGTGTTCACAGATTTAAAGCTTGTGTTTCTAATGAGAGAGGTTCCATGATTAAGACAA | 1560 |
| QY | 1561 | ACCGCCAGTCTCCATCAATCAAGATGAGATGAAAATGCCACCTTGGATGGGACTCTCC | 1620 |
| Db | 1561 | ACCGCCAGTCTCCATCAATCAAGATGAGATGAAAATGCCACCTTGGATGGGACTCTCC | 1620 |
| QY | 1621 | CCACTCCGATATCCAGAACTCGCCCAAGCTGACCCCAATGAAAAAAGACAAAGAGGC | 1680 |
| Db | 1621 | CCACTCCGATATCCAGAACTCGCCCAAGCTGACCCCAATGAAAAAAGACAAAGAGGC | 1680 |
| QY | 1681 | TTTCCAGGGGCAAGAAAGAGTGAAGCAGCTGAGGGCATGTGACTGACATAGGGGTGCT | 1740 |
| Db | 1681 | TTTCCAGGGGCAAGAAAGAGTGAAGCAGCTGAGGGCATGTGACTGACATAGGGGTGCT | 1740 |
| QY | 1741 | GGCAGAGCAAGAAAGGCCACCTCTCTCTGAGACAGTGAACAGGGGCCCACTCCGAAAGAG | 1800 |
| Db | 1741 | GGCAGAGCAAGAAAGGCCACCTCTCTCTGAGACAGTGAACAGGGGCCCACTCCGAAAGAG | 1800 |
| QY | 1801 | AGAGGCAAGCAGATCCACTGTGGCCACCTCGCTTACAGAGGACATCTGCACAGCATGGA | 1860 |
| Db | 1801 | AGAGGCAAGCAGATCCACTGTGGCCACCTCGCTTACAGAGGACATCTGCACAGCATGGA | 1860 |
| QY | 1861 | TCTGAGATATCAAGAGGGTAAACTGTTTGGAAATCGCGCAGTGTGGGAAGTGGAAAAAC | 1920 |
| Db | 1861 | TCTGAGATATCAAGAGGGTAAACTGTTTGGAAATCGCGCAGTGTGGGAAGTGGAAAAAC | 1920 |
| QY | 1921 | CTCTCTCATTTTCAGCCATTTTAAAGCCAGATGACGCTTACAGGGCAGCATTTGCATACG | 1980 |
| Db | 1921 | CTCTCTCATTTTCAGCCATTTTAAAGCCAGATGACGCTTACAGGGCAGCATTTGCATACG | 1980 |
| QY | 1981 | TGGAACCTTCGCTTATGTGGCCACAGGCGCTGTGATCTCTCAATGCTACTCTGAGAGCAA | 2040 |
| Db | 1981 | TGGAACCTTCGCTTATGTGGCCACAGGCGCTGTGATCTCTCAATGCTACTCTGAGAGCAA | 2040 |
| QY | 2041 | CATCCTGTTTGGGAAGAAATGATGAAGAAAGATACAACTGTGCTGAACAGCTGCTG | 2100 |
| Db | 2041 | CATCCTGTTTGGGAAGAAATGATGAAGAAAGATACAACTGTGCTGAACAGCTGCTG | 2100 |
| QY | 2101 | CGTGAGGCTCAGCTGGGCAATTTCTCCAGCAGCGACCTGAAGGAGATTGGAGAGCGAGG | 2160 |
| Db | 2101 | CGTGAGGCTCAGCTGGGCAATTTCTCCAGCAGCGACCTGAAGGAGATTGGAGAGCGAGG | 2160 |
| QY | 2161 | AGCCAACTGAGCGGTGGGCAAGCGCAAGAGATACGCTTGTCCGGGCGCTTGTATAGTGA | 2220 |
| Db | 2161 | AGCCAACTGAGCGGTGGGCAAGCGCAAGAGATACGCTTGTCCGGGCGCTTGTATAGTGA | 2220 |

| | | | |
|----|------|--|------|
| QY | 2221 | CAGAGCATCTCAATCCTTGSAGACACCCCTCACTGCTTAAATGCGCCATGTTGGGCACCA | 2280 |
| Db | 2221 | CAGAGCATCTCAATCCTTGSAGACACCCCTCACTGCTTAAATGCGCCATGTTGGGCACCA | 2280 |
| QY | 2281 | CATCTTCAATAGTGCATTCGGGAAACATCTCAAGTCCAAAGACAGTTCGTTGTTACCCA | 2340 |
| Db | 2281 | CATCTTCAATAGTGCATTCGGGAAACATCTCAAGTCCAAAGACAGTTCGTTGTTACCCA | 2340 |
| QY | 2341 | CCAGTTCAGTACCTGGTTGCTAGTATGATGATCTTCATCAGAAAGAGGCGTGTATTAC | 2400 |
| Db | 2341 | CCAGTTCAGTACCTGGTTGCTAGTATGATGATCTTCATCAGAAAGAGGCGTGTATTAC | 2400 |
| QY | 2401 | GGAAGAAGCACCCTATGAGAACTGATGAAATTTAAATGCTGACTTGTACCATTTTAA | 2460 |
| Db | 2401 | GGAAGAAGCACCCTATGAGAACTGATGAAATTTAAATGCTGACTTGTACCATTTTAA | 2460 |
| QY | 2461 | TAACTTGTGTGGAGAGACACCGGCAGTTGAGATTCATTTAAAAAAAGAACACAGGG | 2520 |
| Db | 2461 | TAACTTGTGTGGAGAGACACCGGCAGTTGAGATTCATTTAAAAAAAGAACACAGGG | 2520 |
| QY | 2521 | TTTCACAGAAGAGTCAACAAAGACAGGCTCTTAAACAGATCAATTAAGAAAGAAAAAGC | 2580 |
| Db | 2521 | TTTCACAGAAGAGTCAACAAAGACAGGCTCTTAAACAGATCAATTAAGAAAGAAAAAGC | 2580 |
| QY | 2581 | AGTAAAGCCAGAGAGAGGAGCGACTTGTGACGTGGAAGAAAGGGCAGGTTCACTGTC | 2640 |
| Db | 2581 | AGTAAAGCCAGAGAGAGGAGCGACTTGTGACGTGGAAGAAAGGGCAGGTTCACTGTC | 2640 |
| QY | 2641 | CTGGTCAGTTATGATGCTCACTACATCCAGGCTGTGGGGGCGCTTGGGATTCGTGGTAT | 2700 |
| Db | 2641 | CTGGTCAGTTATGATGCTCACTACATCCAGGCTGTGGGGGCGCTTGGGATTCGTGGTAT | 2700 |
| QY | 2701 | TATGGCCCTTTTCATGCTGAATGTAGGACAGACCGGCTTCACACACTGGTGGTGAAGTTA | 2760 |
| Db | 2701 | TATGGCCCTTTTCATGCTGAATGTAGGACAGACCGGCTTCACACACTGGTGGTGAAGTTA | 2760 |
| QY | 2761 | CTGATCAAGCAGAGAGCGGGAACACACTGTGACTCGAGGAAACAGACTCGGTAG | 2820 |
| Db | 2761 | CTGATCAAGCAGAGAGCGGGAACACACTGTGACTCGAGGAAACAGACTCGGTAG | 2820 |
| QY | 2821 | TGACAGCATGAGAGACATCTCATATGACAGTATGTCACACTCTACGCCCTCCAT | 2880 |
| Db | 2821 | TGACAGCATGAGAGACATCTCATATGACAGTATGTCACACTCTACGCCCTCCAT | 2880 |
| QY | 2881 | GGCAGTCATGCTGATCTCTGAAAGCCATTTGAGAGAGTTGCTTTGTCAAGGGCAGCTGCG | 2940 |
| Db | 2881 | GGCAGTCATGCTGATCTCTGAAAGCCATTTGAGAGAGTTGCTTTGTCAAGGGCAGCTGCG | 2940 |
| QY | 2941 | AGCTTCTCCCGGCTGATAGCAGAGCTTTTCCGAAAGATCTTTCAAGACCTATGAAGTT | 3000 |
| Db | 2941 | AGCTTCTCCCGGCTGATAGCAGAGCTTTTCCGAAAGATCTTTCAAGACCTATGAAGTT | 3000 |
| QY | 3001 | TTTTGACACGACCCCCCAGGAGGAAFTTCCAAAGGTTTCCAAAGACATGAGTGAAGT | 3060 |
| Db | 3001 | TTTTGACACGACCCCCCAGGAGGAAFTTCCAAAGGTTTCCAAAGACATGAGTGAAGT | 3060 |
| QY | 3061 | TGACGTGCGGCTGCGCTTCAGGCGCAGATGTTCAATCAGAAAGTTCATCTGTTGTT | 3120 |
| Db | 3061 | TGACGTGCGGCTGCGCTTCAGGCGCAGATGTTCAATCAGAAAGTTCATCTGTTGTT | 3120 |
| QY | 3121 | CTGTGTGGGAATGATCGCAGAGATCTTCCCGTGGTCTTGTGGCAGTGGGGCCCTTGT | 3180 |
| Db | 3121 | CTGTGTGGGAATGATCGCAGAGATCTTCCCGTGGTCTTGTGGCAGTGGGGCCCTTGT | 3180 |
| QY | 3181 | CATCTCTTTTCACTCTGTGACATTTGTCTCAGGGTCTGATTCGGGAGCTGAAGCTCT | 3240 |
| Db | 3181 | CATCTCTTTTCACTCTGTGACATTTGTCTCAGGGTCTGATTCGGGAGCTGAAGCTCT | 3240 |
| QY | 3241 | GGACAAATTCACGAGACACCTTTCCTCCTCCCAATCAGTCCAGCATTAACAGGGCTTGC | 3300 |
| Db | 3241 | GGACAAATTCACGAGACACCTTTCCTCCTCCCAATCAGTCCAGCATTAACAGGGCTTGC | 3300 |
| QY | 3301 | CACCATCCAGCCCTCAATTAAGGGCAGAGTTCCTGCACAGATACAGAGACTGCTGGA | 3360 |

| | | | |
|----|------|--|------|
| Db | 3301 | CACCATCCAGCCCTACATATAAGGGCAGGGATTCTGCACAGATATCCAGAGCTCTGTGA | 3360 |
| QY | 3361 | TGACAAACCAAGCTCTTTTATTTTGTATTTACGTGTGCGATGCGGTGGCTGTGCGGCT | 3420 |
| Db | 3361 | TGACAAACCAAGCTCTTTTATTTTGTATTTACGTGTGCGATGCGGTGGCTGTGCGGCT | 3420 |
| QY | 3421 | GGACCTCATCAGATCGCCCTCATCACACACAGGGGGTGTATGATCTTATATCAGCG | 3480 |
| Db | 3421 | GGACCTCATCAGATCGCCCTCATCACACACAGGGGGTGTATGATCTTATATCAGCG | 3480 |
| QY | 3481 | GCAATTTCCCCACACCTATGCGGGTCTCGCATCTCTTATGCTGTCCAGTTAAAGGGGCT | 3540 |
| Db | 3481 | GCAGATTTCCCCACACCTATGCGGGTCTCGCATCTCTTATGCTGTCCAGTTAAAGGGGCT | 3540 |
| QY | 3541 | GTTCCAGTTTACGGTTCAGACTGGCATCTGAGACAGAAAGTCCATTACCTCGTGGGAAG | 3600 |
| Db | 3541 | GTTCCAGTTTACGGTTCAGACTGGCATCTGAGACAGAAAGTCCATTACCTCGTGGGAAG | 3600 |
| QY | 3601 | GATCAATCACTACTATTAAAGCTGTCTCTTGAAGCAGCTGCCAGAAATTAAACAAAGGC | 3660 |
| Db | 3601 | GATCAATCACTACTATTAAAGCTGTCTCTTGAAGCAGCTGCCAGAAATTAAACAAAGGC | 3660 |
| QY | 3661 | TCCCTCCCTACACTGCGCCCAAGAGGAGAGGTGACCTTTGAAAGCGAGATAGGTA | 3720 |
| Db | 3661 | TCCCTCCCTACACTGCGCCCAAGAGGAGAGGTGACCTTTGAAAGCGAGATAGGTA | 3720 |
| QY | 3721 | CCGAGAAAACCTCCCTCGCTCCAAAAGAAATGCTTACAGATCAAACTPAAAGAGA | 3780 |
| Db | 3721 | CCGAGAAAACCTCCCTCGCTCCAAAAGAAATGCTTACAGATCAAACTPAAAGAGA | 3780 |
| QY | 3781 | GATTGGCAATTGTGGGGCGAGAGATCAGGAAGTCTCTGCTGGGGATGCCCCCTTCCG | 3840 |
| Db | 3781 | GATTGGCAATTGTGGGGCGAGAGATCAGGAAGTCTCTGCTGGGGATGCCCCCTTCCG | 3840 |
| QY | 3841 | TCTGTGAGATTTATCTGAGAGCTGCATCAAGATTTGATGGAAGTGAAGATACAGATATGG | 3900 |
| Db | 3841 | TCTGTGAGATTTATCTGAGAGCTGCATCAAGATTTGATGGAAGTGAAGATACAGATATGG | 3900 |
| QY | 3901 | CTTTGCCACCTCCGAGCAAAACCTCTATCATTTCTCTCAAGAGCGGTCTTTCAGTGG | 3960 |
| Db | 3901 | CTTTGCCACCTCCGAGCAAAACCTCTATCATTTCTCTCAAGAGCGGTCTTTCAGTGG | 3960 |
| QY | 3961 | CACCTGCAGATCAAAATTTTGAACCCCTTCAACACAGTACACTGAAAGCAGATTTGGGATGC | 4020 |
| Db | 3961 | CACCTGCAGATCAAAATTTTGAACCCCTTCAACACAGTACACTGAAAGCAGATTTGGGATGC | 4020 |
| QY | 4021 | CCTGAGAGGACACACATGAAGAATATATTGCTCAGCTACTGTGAACCTTGAATCTGA | 4080 |
| Db | 4021 | CCTGAGAGGACACACATGAAGAATATATTGCTCAGCTACTGTGAACCTTGAATCTGA | 4080 |
| QY | 4081 | AGTAGTAGAAGATGGGATTAATCTTCAGTGGGGGGAACGGCAGCTTGTGATATGCTAG | 4140 |
| Db | 4081 | AGTAGTAGAAGATGGGATTAATCTTCAGTGGGGGGAACGGCAGCTTGTGATATGCTAG | 4140 |
| QY | 4141 | AGCCCTGTCTCCGCACAGTAAAGATTCTGATTTTATGATGAAGCCACAGCTGCCATGGACAC | 4200 |
| Db | 4141 | AGCCCTGTCTCCGCACAGTAAAGATTCTGATTTTATGATGAAGCCACAGCTGCCATGGACAC | 4200 |
| QY | 4201 | AGAGACAGACTTATTTGATTTCAAGAGACCATCCGAGAAGCATTTTGCAGACTGTACATGCT | 4260 |
| Db | 4201 | AGAGACAGACTTATTTGATTTCAAGAGACCATCCGAGAAGCATTTTGCAGACTGTACATGCT | 4260 |
| QY | 4261 | GACCAATTCGCATGGCTCGCACACAGGTTCTAGGCTCGAGTAGGATATATGCTCTGGCCA | 4320 |
| Db | 4261 | GACCAATTCGCATGGCTCGCACACAGGTTCTAGGCTCGAGTAGGATATATGCTCTGGCCA | 4320 |
| QY | 4321 | GGGACAGGTGTGTGATTTGACACCCCATGGTCTCTTGTCCACAGACAGTATCCCGATT | 4380 |
| Db | 4321 | GGGACAGGTGTGTGATTTGACACCCCATGGTCTCTTGTCCACAGACAGTATCCCGATT | 4380 |
| QY | 4381 | CTATCCATATGTTTCTGCTGACAGAAACAAGTCTGCTGTCACAGGCTGACTCTCTCTGT | 4440 |
| Db | 4381 | CTATCCATATGTTTCTGCTGACAGAAACAAGTCTGCTGCTGTCACAGGCTGACTCTCTCTGT | 4440 |

| | | | |
|---|------|---|------|
| Db | 4381 | CTATGCCATGCTTTGGCTGCTGCACAGAAACAAGGTCGCTGCACAGGGGCTGACTCTCTCCCTGT | 4440 |
| QY | 4441 | TGACGAAGCTCTCTTTCTTTTAGAGCAATTGCCATTCCTGCGGAGCGGGCCCTTCATC | 4500 |
| Db | 4441 | TGACGAAGCTCTCTTTCTTTTAGAGCAATTGCCATTCCTGCGGAGCGGGCCCTTCATC | 4500 |
| QY | 4501 | GCGGCCCTCCACGGAACCTTGCCCTTCTGCATTTATCTTCGACACAGCACTCCGAT | 4560 |
| Db | 4501 | GCGGCCCTCCACGGAACCTTGCCCTTCTGCATTTATCTTCGACACAGCACTCCGAT | 4560 |
| QY | 4561 | TGGCTGTGTGTTTCACCTTTTAGGAGAGCTCATATTTGATTATTTGTAATTTCCATAT | 4620 |
| Db | 4561 | TGGCTGTGTGTTTCACCTTTTAGGAGAGCTCATATTTGATTATTTGTAATTTCCATAT | 4620 |
| QY | 4621 | TCAAGTAAACAAAATTTAGTTTTGTCTTAATTCGACTCTAAAGCTCAGGAACCGT | 4680 |
| Db | 4621 | TCAAGTAAACAAAATTTAGTTTTGTCTTAATTCGACTCTAAAGCTCAGGAACCGT | 4680 |
| QY | 4681 | TATATATATTTGTATCAGAGGCTATATATGAAGCTTATACGTGTAGCTATATCATATAT | 4740 |
| Db | 4681 | TATATATATTTGTATCAGAGGCTATATATGAAGCTTATACGTGTAGCTATATCATATAT | 4740 |
| QY | 4741 | AATTCGTACATAGCCATATTTTACGTGAAATGTACACTGTTTATTTTATATTAAT | 4800 |
| Db | 4741 | AATTCGTACATAGCCATATTTTATTTTAAATGCACCTCTTAAAGGTTACAGGAACCGT | 4800 |
| QY | 4801 | AAGCAGCTGCTCTAAAAAAGGCGGCGCCG 4847 | |
| Db | 4801 | AAGCAGCTGCTCTAAAAAAGGCGGCGCCG 4847 | |
| RESULT 2 | | | |
| US-09-001-273-1 | | | |
| : Sequence 1, Application us/09001273 | | | |
| : Patent No. 5994130 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: SHITJAN, Andrew | | | |
| TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED | | | |
| TITLE OF INVENTION: POLYPEPTIDE | | | |
| NUMBER OF SEQUENCES: 8 | | | |
| CORRESPONDENCE ADDRESS: | | | |
| ADDRESSEE: Testa, Hurwitz & Thibault | | | |
| STREET: 125 High St. | | | |
| CITY: Boston | | | |
| STATE: MA | | | |
| COUNTRY: USA | | | |
| ZIP: 02110 | | | |
| COMPUTER READABLE FORM: | | | |
| MEDIUM TYPE: Floppy disk | | | |
| COMPUTER: IBM PC compatible | | | |
| OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| SOFTWARE: PatentIn Release #1.0, Version #1.30 | | | |
| CURRENT APPLICATION DATA: | | | |
| APPLICATION NUMBER: US/09/001,273 | | | |
| FILING DATE: | | | |
| CLASSIFICATION: | | | |
| ATTORNEY/AGENT INFORMATION: | | | |
| NAME: FENTON, Gillian M | | | |
| REGISTRATION NUMBER: 36,508 | | | |
| REFERENCE/DOCKET NUMBER: MIT-001 | | | |
| TELECOMMUNICATION INFORMATION: | | | |
| TELEPHONE: (617) 248-7000 | | | |
| TELEFAX: (617) 248-7100 | | | |
| INFORMATION FOR SEQ. ID NO. 1: | | | |
| SEQUENCE CHARACTERISTICS: | | | |
| LENGTH: 4781 base pairs | | | |
| TYPE: nucleic acid | | | |
| STRANDEDNESS: single | | | |
| TOPOLOGY: linear | | | |
| FEATURE: | | | |
| NAME/KEY: CDS | | | |
| LOCATION: 2..4360 | | | |
| US-09-001-273-1 | | | |

| Query Match | 98.5% | Score 4775.4 | DB 2 | Length 4781 |
|-----------------------|--|--------------|----------|-------------|
| Best Local Similarity | 99.8% | Fred. No. 0 | | |
| Matches 4773 | Conservative 6 | Matches 2 | Indels 0 | Gaps 0 |
| OY | 7GATGTGAAGCAATACAGTCTGTGACCCCTGGAAACCTCAGCTACAGAGAATGAAGATAT | 126 | | |
| Db | 1TGATGTGAAGCAATACAGTCTGTGACCCCTGGAAACCTCAGCTACAGAGAATGAAGATAT | 60 | | |
| OY | 127CGACATAGAAAAGAGTATATCATGCCAGTCTTGGGTATATAGAGTGTGAGGAGAGAAC | 186 | | |
| Db | 61CGACATAGAAAAGAGTATATCATGCCAGTCTTGGGTATATAGAGTGTGAGGAGAGAAC | 120 | | |
| OY | 187CAGCACTTCTGGGAGGACACAGAGCCGGTGAATTCGAAATTCCAGATTCCAGAGAACTCGACCTT | 246 | | |
| Db | 121CAGCACTTCTGGGAGGACACAGAGCCGGTGAATTCGAAATTCCAGATTCCAGAGAACTCGACCTT | 180 | | |
| OY | 247GGAATGCCAAGATGCCCTTGAAAAACAGACGCCGAGCGAGGGCCCTCTCTTGATGCTTC | 306 | | |
| Db | 181GGAATGCCAAGATGCCCTTGAAAAACAGACGCCGAGCGAGGGCCCTCTCTTGATGCTTC | 240 | | |
| OY | 307CATGATTTCTAGCTTCAGATTCCTGGATGAGAGCATCCCAAGGGAAAGTACCATCATAGG | 366 | | |
| Db | 241CATGATTTCTAGCTTCAGATTCCTGGATGAGAGCATCCCAAGGGAAAGTACCATCATAGG | 300 | | |
| OY | 367CTTGAGTCTCTGAGAGCCCATCCGAGCTACTTGTCAAAACACAGACCCAGAGTGAAGATGC | 426 | | |
| Db | 301CTTGAGTCTCTGAGAGCCCATCCGAGCTACTTGTCAAAACACAGACCCAGAGTGAAGATGC | 360 | | |
| OY | 427TGGGCTTTTTCCTGTATATGACTTTTTCGTGGCTTTTCTCTCTGCGCCGTGTGGCCACAA | 486 | | |
| Db | 361TGGGCTTTTTCCTGTATATGACTTTTTCGTGGCTTTTCTCTCTGCGCCGTGTGGCCACAA | 420 | | |
| OY | 487GAAGGGGAGGCTCTCAATGGAAGAGCTGTGGCTCTGTCCAAAGCACAGATCTTCTGAGCT | 546 | | |
| Db | 421GAAGGGGAGGCTCTCAATGGAAGAGCTGTGGCTCTGTCCAAAGCACAGATCTTCTGAGCT | 480 | | |
| OY | 547GAACGTCAAGAAAGACTAGAGAGACTGTGGCAAGAAAGACTGAATGGAAGTTGGGCCAGAGCC | 606 | | |
| Db | 481GAACGTCAAGAAAGACTAGAGAGACTGTGGCAAGAAAGACTGAATGGAAGTTGGGCCAGAGCC | 540 | | |
| OY | 607TGCCTCCCTGGCAAGGGTTGTGTGATCTCTGCGGCAACAGGCTCATCTGATCATGCT | 666 | | |
| Db | 541TGCCTCCCTGGCAAGGGTTGTGTGATCTCTGCGGCAACAGGCTCATCTGATCATGCT | 600 | | |
| OY | 667GTGCTGTATGATCAGCAGCTGGCTGGCTTCAGTGGACCAAGCCTTCATGTGAAACACT | 726 | | |
| Db | 601GTGCTGTATGATCAGCAGCTGGCTGGCTTCAGTGGACCAAGCCTTCATGTGAAACACT | 660 | | |
| OY | 727CTTGGAGTATACCCAGGCAACAGATCTTAACTGTGAGTACAGCTTGTATAGTACTGGG | 786 | | |
| Db | 661CTTGGAGTATACCCAGGCAACAGATCTTAACTGTGAGTACAGCTTGTATAGTACTGGG | 720 | | |
| OY | 787CTTCTCTGTAGAGAAATCTGTGGTCTTGTGCTGTGAGTACAGCTTGTGGCATTAATTA | 846 | | |
| Db | 721CTTCTCTGTAGAGAAATCTGTGGTCTTGTGCTGTGAGTACAGCTTGTGGCATTAATTA | 780 | | |
| OY | 847CCGAACCCGGTGTCCGCTTGGCGGGGGCCATCTTAACCATGSCATTTAAGAGATCCTTAA | 906 | | |
| Db | 781CCGAACCCGGTGTCCGCTTGGCGGGGGCCATCTTAACCATGSCATTTAAGAGATCCTTAA | 840 | | |
| OY | 907GTTAAGAACAATTAAGAAGAAATCCCTGGGTGAGTCATCAACATTTGCTCCAAAGATGG | 966 | | |
| Db | 841GTTAAGAACAATTAAGAAGAAATCCCTGGGTGAGTCATCAACATTTGCTCCAAAGATGG | 900 | | |
| OY | 967GCAGAGAAATGTTTGAAGGACAGCCGTTTGGCAGCCTGTGTGGCTGGAGGACCCGTTGTTGC | 1026 | | |
| Db | 901GCAGAGAAATGTTTGAAGGACAGCCGTTTGGCAGCCTGTGTGGAGGACCCGTTGTTGC | 960 | | |
| OY | 1027CATCTTAGCATGATTTATATATATATATTTCTGGGACCAACAGCCTTCTTGAGTACAGC | 1086 | | |
| Db | 961CATCTTAGCATGATTTATATATATATATATTTCTGGGACCAACAGCCTTCTTGAGTACAGC | 1020 | | |

| | | | |
|----|------|--|------|
| Qy | 1087 | TGTTTTATCTCTTTTACCAGCAATGATGTTTGCATCAGGGCTCACAGCATATTTAC | 1146 |
| Dp | 1021 | TGTTTTATCTCTTTTACCAGCAATGATGTTTGCATCAGGGCTCACAGCATATTTAC | 1080 |
| Qy | 1147 | GAGAAATGCGTGGCCGCCACGGATACAGTGTGCCAGAAAGATGAAGTGAAGTCTTACTTA | 1206 |
| Dp | 1081 | GAGAAATGCGTGGCCGCCACGGATACAGTGTGCCAGAAAGATGAAGTGAAGTCTTACTTA | 1140 |
| Qy | 1207 | CATTAAATTTATCAAAATGTATGCTGGGTCAAAAGCATTTTCTCAGGTGTTCAGAAAT | 1266 |
| Dp | 1141 | CATTAAATTTATCAAAATGTATGCTGGGTCAAAAGCATTTTCTCAGGTGTTCAGAAAT | 1200 |
| Qy | 1267 | CCGGAGAGAGACGTCGGAATATGGAAAAACCCGGGTAATCCAGACATACATCTGGGG | 1326 |
| Dp | 1201 | CCGGAGAGAGACGTCGGAATATGGAAAAACCCGGGTAATCCAGACATACATCTGGGG | 1260 |
| Qy | 1327 | TGTGGCTCCCATTTGTGTGTGTGTAATGGCCAGCGTGGTGACCTTCTCTGTTCATATGACCT | 1386 |
| Dp | 1261 | TGTGGCTCCCATTTGTGTGTGTGTAATGGCCAGCGTGGTGACCTTCTCTGTTCATATGACCT | 1320 |
| Qy | 1387 | GGGCTTGCATCTGACACGACGACACAGGCTTTCACAGTGTGACAGTCTTCAATTCATAC | 1446 |
| Dp | 1321 | GGGCTTGCATCTGACACGACGACACAGGCTTTCACAGTGTGACAGTCTTCAATTCATAC | 1380 |
| Qy | 1447 | TTTTGGCTTTGAAGTAAACCCGTTTTCACTTAAAGTCCCTCAGAAAGCTCAGTGGCGT | 1506 |
| Dp | 1381 | TTTTGGCTTTGAAGTAAACCCGTTTTCACTTAAAGTCCCTCAGAAAGCTCAGTGGCGT | 1440 |
| Qy | 1507 | TGACAGATTTAAAGATTTGTTTCTAATGGAAGAGGTTCCATGATAAAGAACAAACGAC | 1566 |
| Dp | 1441 | TGACAGATTTAAAGATTTGTTTCTAATGGAAGAGGTTCCATGATAAAGAACAAACGAC | 1500 |
| Qy | 1567 | CAGTCCACATCAAGATGAGATGSAAAAAATGCACCTTGGATGGGACTCCTCCCACTC | 1626 |
| Dp | 1501 | CAGTCCACATCAAGATGAGATGSAAAAAATGCACCTTGGATGGGACTCCTCCCACTC | 1560 |
| Qy | 1627 | CAGTATCCAGACTCGCCCAAGCTGACCCCCAAAATGAAAAAAGCAAGAGGGCTTCAG | 1686 |
| Dp | 1561 | CAGTATCCAGACTCGCCCAAGCTGACCCCCAAAATGAAAAAAGCAAGAGGGCTTCAG | 1620 |
| Qy | 1687 | GGGCAAAAGAGAGAGGTGAGGAGCTGAGAGGCACTGACATCGAGGGGTGTGGCAGA | 1746 |
| Dp | 1621 | GGGCAAAAGAGAGAGGTGAGGAGCTGAGAGGCACTGACATCGAGGGGTGTGGCAGA | 1680 |
| Qy | 1747 | GCAGAAAGGGCCACTCCTCTGTGACAGTACAGACAGAGGGCCACTCCCGAAGAGAAAGG | 1806 |
| Dp | 1681 | GCAGAAAGGGCCACTCCTCTGTGACAGTACAGAGAGGGCCACTCCCGAAGAGAAAGG | 1740 |
| Qy | 1807 | CAGACATCTCACCCTGGGCCACTGGCGTTACAGAGACACTGCACAGCATCAGTCTGA | 1866 |
| Dp | 1741 | CAGACATCTCACCCTGGGCCACTGGCGTTACAGAGACACTGCACAGCATCAGTCTGA | 1800 |
| Qy | 1867 | GATCAAGAGGGTAACTGGTTGGATCTMGCGGCAAGTGGGGAAGTGGAAAAACCTCTCT | 1926 |
| Dp | 1801 | GATCAAGAGGGTAACTGGTTGGATCTMGCGGCAAGTGGGGAAGTGGAAAAACCTCTCT | 1860 |
| Qy | 1927 | CATTTCAGCCATTTTAGCCAGATGACGCTTTCAGAGGGCAGACATTCGATCAGTGGAC | 1986 |
| Dp | 1861 | CATTTCAGCCATTTTAGCCAGATGACGCTTTCAGAGGGCAGACATTCGATCAGTGGAC | 1920 |
| Qy | 1987 | CTTGCGTTATGTGGCCACGACGGCTTGGATCCTCAATGCTACTCTGAGAGACAACTCTCT | 2046 |
| Dp | 1921 | CTTGCGTTATGTGGCCACGACGGCTTGGATCCTCAATGCTACTCTGAGAGACAACTCTCT | 1980 |
| Qy | 2047 | GTTTGGGAAGGAATATGATTAAGAAAGATACAACCTGTGCTGAGACACTCTGCTGCTAG | 2106 |
| Dp | 1981 | GTTTGGGAAGGAATATGATTAAGAAAGATACAACCTGTGCTGAGACACTCTGCTGCTAG | 2040 |
| Qy | 2107 | GCTGACCTTGCCATCTTCTCCACGACGACCTGACGAGATTTGAGAGACGAGAGCCAA | 2166 |
| Dp | 2041 | GCTGACCTTGCCATCTTCTCCACGACGACCTGACGAGATTTGAGAGACGAGAGCCAA | 2100 |

| | | | |
|----|------|--|------|
| QY | 2167 | CCTGAGGGGTGGGACGCGCCAGAGGATCAAGCTTGGCCGGGCTGTATAGTACAGAG | 2228 |
| Db | 2101 | CCTGAGGGGTGGGACGCGCCAGAGGATCAAGCTTGGCCGGGCTGTATAGTACAGAG | 2160 |
| QY | 2227 | CATCATCTCTGGAGCAACCCCTCAGTGCGTTAGATGGCCATGTGGGCAACCACTATT | 2286 |
| Db | 2161 | CATCATCTCTGGAGCAACCCCTCAGTGCGTTAGATGGCCATGTGGGCAACCACTATT | 2220 |
| QY | 2287 | CAATAGTCTATCCGGAAACATCTCAAGTCCAAAGACAGTTCGTTGTTTACCACAGTT | 2346 |
| Db | 2221 | CAATAGTCTATCCGGAAACATCTCAAGTCCAAAGACAGTTCGTTGTTTACCACAGTT | 2280 |
| QY | 2347 | ACAGTACCTGGTTCAGTGTGATGAAGTATCTTCATGAAGAAGGCGTGTATTACGAAAG | 2406 |
| Db | 2281 | ACAGTACCTGGTTCAGTGTGATGAAGTATCTTCATGAAGAAGGCGTGTATTACGAAAG | 2340 |
| QY | 2407 | AGGCAACCATGAGCACTGATGAATTTAAATGTTGCTGTACTACATTTTAAATAACT | 2466 |
| Db | 2341 | AGGCAACCATGAGCACTGATGAATTTAAATGTTGCTGTACTACATTTTAAATAACT | 2400 |
| QY | 2467 | GTTCGTGGGAGAGACACCGCAGTTGAGATCAATTAATAAAGAAACAGTGTTTACA | 2566 |
| Db | 2401 | GTTCGTGGGAGAGACACCGCAGTTGAGATCAATTAATAAAGAAACAGTGTTTACA | 2460 |
| QY | 2527 | GAAGAAGTCACAAAGACAAAGGTCCTAAACAGGATCAATTAAGAAAGAAAAAGCAGTAA | 2586 |
| Db | 2461 | GAAGAAGTCACAAAGAGGTCCTAAACAGGATCAATTAAGAAAGAAAAAGCAGTAA | 2520 |
| QY | 2587 | GCCAGAGGAAGGCGACTTGTGCAGCTGGAAAGAAAGGCGCAGGGTTCAGTCCCTGGTC | 2646 |
| Db | 2521 | GCCAGAGGAAGGCGACTTGTGCAGCTGGAAAGAAAGGCGCAGGGTTCAGTCCCTGGTC | 2580 |
| QY | 2647 | AGTATATGTCTTACATCCAGCGCTGCGGGGCCCTTGGCATTCCTGTGTTATTATGCG | 2706 |
| Db | 2581 | AGTATATGTCTTACATCCAGCGCTGCGGGGCCCTTGGCATTCCTGTGTTATTATGCG | 2640 |
| QY | 2707 | CTTTTCATGCTGATGTAGGACAGACCGCCTTCAGCACCTGTGTGTGATTTACTGAT | 2766 |
| Db | 2641 | CTTTTCATGCTGATGTAGGACAGACCGCCTTCAGCACCTGTGTGTGATTTACTGAT | 2700 |
| QY | 2767 | CAAGCAAGGAAGGGGAAACCACTGTGACTGAGGGAAGAGACCTGGTGTAGTACAGAG | 2826 |
| Db | 2701 | CAAGCAAGGAAGGGGAAACCACTGTGACTGAGGGAAGAGACCTGGTGTAGTACAGAG | 2760 |
| QY | 2827 | CATGAAGACAAATCTCATATATGACAGTACTATGTCAGACATCTACGCCCTCCATGGCAGT | 2886 |
| Db | 2761 | CATGAAGACAAATCTCATATATGACAGTACTATGTCAGACATCTACGCCCTCCATGGCAGT | 2820 |
| QY | 2887 | CATGCTGATCTGAAAGCCATTCGAGAGATTGTCTTTGTCAAAGGCGACGTCGAGCTTC | 2946 |
| Db | 2821 | CATGCTGATCTGAAAGCCATTCGAGAGATTGTCTTTGTCAAAGGCGACGTCGAGCTTC | 2880 |
| QY | 2947 | CTCCCGGCTGCATGACGAGCTTTTCCGAAAGGATCTTCCAGAGCCTATGAAAGTTTTTTGA | 3006 |
| Db | 2881 | CTCCCGGCTGCATGACGAGCTTTTCCGAAAGGATCTTCCAGAGCCTATGAAAGTTTTTTGA | 2940 |
| QY | 3007 | CACGACCCCAACAGGAGGAAATTCOAACAGGTTTCCAAAGAACATGAAAGTAAGTGAAGT | 3066 |
| Db | 2941 | CACGACCCCAACAGGAGGAAATTCOAACAGGTTTCCAAAGAACATGAAAGTAAGTGAAGT | 3000 |
| QY | 3067 | GGCGCTGCCGTTCCAGGCGCAAGATGTTTCATCCAGAAAGTTATCTGTGTTCTCTGTGT | 3126 |
| Db | 3001 | GGCGCTGCCGTTCCAGGCGCAAGATGTTTCATCCAGAAAGTTATCTGTGTTCTCTGTGT | 3060 |
| QY | 3127 | GGGAATGATGCGAGAGTCTTCCCGTGGTTCTTGTGGCAGTGGGGCCCTTGTATCTCT | 3186 |
| Db | 3061 | GGGAATGATGCGAGAGTCTTCCCGTGGTTCTTGTGGCAGTGGGGCCCTTGTATCTCT | 3120 |
| QY | 3187 | CTTTTCAGTCTCTGCACATTTCTTCACAGGGGTCGATTCGGGAGCTGAAAGCGTCTGGAACA | 3246 |
| Db | 3121 | CTTTTCAGTCTCTGCACATTTCTTCACAGGGGTCGATTCGGGAGCTGAAAGCGTCTGGAACA | 3180 |
| QY | 3247 | TATACGCAATCACCTTCTCTCCACATATCAAGTCCACACATACAGGGCTTTCGCCACAT | 3306 |

| | | | |
|---|------|---|-------|
| D | 3181 | TATCAGCAGTCAACCTTTCTCTCTCCACATCAAGTCCAGCATACAGGGCCTTGCCACCAT | 32448 |
| Q | 3307 | CCAGCCTAACAAATTAAGGGAGAGATTTCGACAGATACCAAGAGCTGCTGGATGACAA | 33668 |
| D | 3341 | CCAGGCTCAAAATAAGGGAGAGATTTCGACAAATACCAAGAGCTGCTGGATGACAA | 33008 |
| Q | 3367 | CCAAAGCTCTTTTCTTTTCTGTTTACGTCGATGCGGTGGCTGGCTGGAGCT | 34288 |
| D | 3301 | CCAAAGCTCTTTTCTTTTCTGTTTACGTCGATGCGGTGGCTGGCTGGAGCT | 33668 |
| Q | 3427 | CATCAGCATGCGCCCTCATCACACCAAGGGGCTGATGATGTTCTTATGACAGGGCAGAT | 34888 |
| D | 3361 | CATCAGCATGCGCCCTCATCACACCAAGGGGCTGATGATGTTCTTATGACAGGGCAGAT | 34208 |
| Q | 3487 | TCCCCAGCCTATGCGGGTCTGGCCATCTTATGCTGACGTTAAAGGGGCTTTGCCA | 35448 |
| D | 3421 | TCCCCAGCCTATGCGGGTCTGGCCATCTTATGCTGACGTTAAAGGGGCTTTGCCA | 34808 |
| Q | 3547 | GTTTACCGTCCAGACTGGCATCTAGACAGAAAGCTCATTCACCTCGGTGGAAGATCAA | 36088 |
| D | 3481 | GTTTACCGTCCAGACTGGCATCTAGACAGAAAGCTCATTCACCTCGGTGGAAGATCAA | 35448 |
| Q | 3607 | TCACATCATTTAAGACTCTGTCTTGGAGACCTGCCAGATTTAAGAACAGGCTCCCTC | 36668 |
| D | 3541 | TCACATCATTTAAGACTCTGTCTTGGAGACCTGCCAGATTTAAGAACAGGCTCCCTC | 36008 |
| Q | 3667 | CCCTGATGCGCCACAGAGGGAGAGGTGACCTTTGAACGCGAGAGATGAGTTACCGGA | 37268 |
| D | 3601 | CCCTGATGCGCCACAGAGGGAGAGGTGACCTTTGAACGCGAGAGATGAGTTACCGGA | 36668 |
| Q | 3727 | AAACCTCCCTCTGCTCTTAAAGAAATATCTTACGATCAAAACCTAAAGAGAAATGG | 37868 |
| D | 3661 | AAACCTCCCTCTGCTCTTAAAGAAATATCTTACGATCAAAACCTAAAGAGAAATGG | 37208 |
| Q | 3787 | CATTGTGGGGGGAGACAGGATCAGAGAAAGTCTGCGGGGATAGCGCTCTCCGCTGGT | 38448 |
| D | 3721 | CATTGTGGGGGGAGACAGGATCAGAGAAAGTCTGCGGGGATAGCGCTCTCCGCTGGT | 37808 |
| Q | 3847 | GGAGTTATCTGAGGCTGCATCAAGATTGATGGAGTGAAGATCAGTGATTTGGCCTTGC | 39068 |
| D | 3781 | GGAGTTATCTGAGGCTGCATCAAGATTGATGGAGTGAAGATCAGTGATTTGGCCTTGC | 38408 |
| Q | 3907 | CGACCTCCGAGCAAACTCTCTATCTCATTTCTCAAGACCGGTGCTGTTCAGTGGCACTGT | 39668 |
| D | 3841 | CGACCTCCGAGCAAACTCTCTATCTCATTTCTCAAGACCGGTGCTGTTCAGTGGCACTGT | 39008 |
| Q | 3967 | CAGATCAAAATTTGGACCCCTTCAACACAGTACACTGAAGACAGATTTGGATGGCCCTGGA | 40268 |
| D | 3901 | CAGATCAAAATTTGGACCCCTTCAACACAGTACACTGAAGACAGATTTGGATGGCCCTGGA | 39668 |
| Q | 4027 | GAGGACACACATGAAGAATGTTTGTCTACGCTACTCTGAAACTTGAATCTGAAGTGAAT | 40868 |
| D | 3961 | GAGGACACACATGAAGAATGTTTGTCTACGCTACTCTGAAACTTGAATCTGAAGTGAAT | 40208 |
| Q | 4087 | GGAGAAATGGGGATTAACCTTCATGTGGGGAGACGGCAGCTTGTGATAGTGAAGACCTT | 41468 |
| D | 4021 | GGAGAAATGGGGATTAACCTTCATGTGGGGAGACGGCAGCTTGTGATAGTGAAGACCTT | 40868 |
| Q | 4147 | GCTTCGCGCACTGTAAGATTCTGATTTTGAATGAAGCCACAGCTGCCATGACACAGAGAC | 42068 |
| D | 4081 | GCTTCGCGCACTGTAAGATTCTGATTTTGAATGAAGCCACAGCTGCCATGACACAGAGAC | 41408 |
| Q | 4207 | AGACTTTTATTTCAAGAGACCAATCCGAGAGAGATTTGAGACTGTACCATCTGACAT | 42668 |
| D | 4141 | AGACTTTTATTTCAAGAGACCAATCCGAGAGAGATTTGAGACTGTACCATCTGACAT | 42008 |
| Q | 4267 | TGCCCATGCGCTCCACACGGTTTCTAGGCTCCGATAGATTAATGTCGTGGCCACAGGACA | 43268 |
| D | 4201 | TGCCCATGCGCTCCACACGGTTTCTAGGCTCCGATAGATTAATGTCGTGGCCACAGGACA | 42668 |
| Q | 4327 | GGTGGTGAAGTTGACACCCCATCGGTCTTCTGTCCAAGCAAGTTCCCATTTTATGC | 43868 |

| | | | |
|---|------|--|------|
| Db | 4261 | GGNNGTGGAGATTGACACCCCATGGCTCTGTCGTCACAGCAGATGCCGATCTGTATGCG | 4320 |
| Qy | 4387 | CAIGTTTGGCTGCTGCAGAGAAACAAGTCCGCTGTCAAAGGCTACTCTCTCCGTGTGACGA | 4446 |
| Db | 4321 | CATGTGTGGCTGCTGCAGAGAAACAAGTCCGCTGTCAAAGGCTACTCTCTCCGTGTGACGA | 4380 |
| Qy | 4447 | AGTCTCTTTCTTTAGAGCATTCGCCAATTCGCCCTGGGGGGGGCCCTTCATCGCGTCC | 4506 |
| Db | 4381 | AGTCTCTTTCTTTAGAGCATTCGCCAATTCGCCCTGGGGGGGGCCCTTCATCGCGTCC | 4440 |
| Qy | 4507 | TCCATCCGAACCTTGCCCTTCTCGATTTTATCTTTCCGACAGCAGTTCCGGATTGGCTT | 4566 |
| Db | 4441 | TCCATCCGAACCTTGCCCTTCTCGATTTTATCTTTCCGACAGCAGTTCCGGATTGGCTT | 4500 |
| Qy | 4567 | GGTGTGTTCACTTTTAAAGGAGAGTCATATTTGGATTATTTGATTTATTCATATTCATGT | 4626 |
| Db | 4501 | GGTGTGTTCACTTTTAAAGGAGAGTCATATTTGGATTATTTGATTTATTCATATTCATGT | 4560 |
| Qy | 4627 | AAACAAATTTAGTTTTTGTCTTATTTTCACCTCTTAAAGTTCAAGGAAACCGTATATAT | 4686 |
| Db | 4561 | AAACAAATTTAGTTTTTGTCTTATTTTCACCTCTTAAAGTTCAAGGAAACCGTATATAT | 4620 |
| Qy | 4687 | AATGTATTCAGAGCCCTAATATGACCTTTATACGTATATTCATATATATATCT | 4746 |
| Db | 4621 | AATGTATTCAGAGCCCTAATATGACCTTTATACGTATATTCATATATATATCT | 4680 |
| Qy | 4747 | GTACATAGCCTATATTTTACAGTGAATAATGTAAAGCTGTTATTTTATATTAATAATAGCAC | 4806 |
| Db | 4681 | GTACATAGCCTATATTTTACAGTGAATAATGTAAAGCTGTTATTTTATATTAATAATAGCAC | 4740 |
| Qy | 4807 | TGTGCTTAAAAAATTT | 4847 |
| Db | 4741 | TGTGCTTAAAAAATTT | 4781 |
| RESULT 3 | | | |
| US-08-843-459A-1 | | | |
| : Sequence 1, Application US/08843459A | | | |
| : Patent No. 6162616 | | | |
| : GENERAL INFORMATION: | | | |
| : APPLICANT: SHYJAN, Andrew | | | |
| : TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED | | | |
| : NUMBER OF SEQUENCES: 8 | | | |
| : CORRESPONDENCE ADDRESS: | | | |
| : ADDRESSEE: LAHIVE & COCKFIELD, LLP | | | |
| : STREET: 28 State Street | | | |
| : City: Boston | | | |
| : STATE: Massachusetts | | | |
| : COUNTRY: USA | | | |
| : Zip: 02109 | | | |
| : COMPUTER READABLE FORM: | | | |
| : MEDIUM TYPE: floppy disk | | | |
| : COMPUTER: IBM PC compatible | | | |
| : OPERATING SYSTEM: PC-DOS/MS-DOS | | | |
| : SOFTWARE: Patent Release #1.0, Version #1.30 | | | |
| : CURRENT APPLICATION DATA: | | | |
| : APPLICATION NUMBER: US/08/843,459A | | | |
| : FILING DATE: 16-Apr-1997 | | | |
| : CLASSIFICATION: 536 | | | |
| : ATTORNEY/AGENT INFORMATION: | | | |
| : NAME: Hanley, Elizabeth A. | | | |
| : REGISTRATION NUMBER: 33,505 | | | |
| : REFERENCE/DOCKET NUMBER: MNI-056 (formerly MIL-001) | | | |
| : TELECOMMUNICATION INFORMATION: | | | |
| : TELEPHONE: (617)227-7400 | | | |
| : TELEFAX: (617)42-4214 | | | |
| : INFORMATION FOR SEQ ID NO: 1: | | | |
| : SEQUENCE CHARACTERISTICS: | | | |
| : LENGTH: 4781 base pairs | | | |
| : type: nucleic acid | | | |
| : STRANDEDNESS: single | | | |
| : TOPOLOGY: linear | | | |

FEATURE:
NAME/KEY: CDS
LOCATION: 2..4360
US-08-843-459A-1

Query Match 98.5%; Score 4775.4; DB 4; Length 4781;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 4773; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 67 TGAATGAAACATACAGTCTGTAGCCCTGGAACCTCCTCAGAGAGATGAAGATAT 126
DB 1 TGAATGAAACATACAGTCTGTAGCCCTGGAACCTCCTCAGAGAGATGAAGATAT 60
QY 127 CCACATAGGAAAAGAT 186
DB 61 CCACATAGGAAAAGAT 120
QY 187 CAGCATCTGGGAGGACAGAGAGCCGAGTCCAGATTCCAGGAGACCTGACCTT 246
DB 121 CAGCATCTGGGAGGACAGAGAGCCGAGTCCAGATTCCAGGAGACCTGACCTT 180
QY 247 GGAATGCCAAGATGCCCTTGGAAACAGACCCGAGCCGAGGCTCTCTCTGATGCTC 306
DB 181 GGAATGCCAAGATGCCCTTGGAAACAGACCCGAGCCGAGGCTCTCTCTGATGCTC 240
QY 307 CATGCATCTCAGCTCAGATCCCTGGATGAGAGCATCCCAAGGAAATACCATCATG 366
DB 241 CATGCATCTCAGCTCAGATCCCTGGATGAGAGCATCCCAAGGAAATACCATCATG 300
QY 367 CTTGATGCTCGAAGCCCATCCGAGCTACTTGCAGAACCCAGCCAGCCAGCAATG 426
DB 301 CTTGATGCTCGAAGCCCATCCGAGCTACTTGCAGAACCCAGCCAGCCAGCAATG 360
QY 427 TGGGCTTTTCTCTGATGACTTTTCTGCTGCTTCTCTGCTGCTGCTGCTGCTGCT 486
DB 361 TGGGCTTTTCTCTGATGACTTTTCTGCTGCTTCTCTGCTGCTGCTGCTGCTGCT 420
QY 487 GAAAGGGAGCTCTCAATGGAAGAGCTGTGCTCTGCTCAAGACAGAGCTTTTGACGT 546
DB 421 GAAAGGGAGCTCTCAATGGAAGAGCTGTGCTCTGCTCAAGACAGAGCTTTTGACGT 480
QY 547 GAACATGACAGACTGAGAGACTGTGCGAAGAGAGCTGAATGAAGTTGGGCGACAGCG 606
DB 481 GAACATGACAGACTGAGAGACTGTGCGAAGAGAGCTGAATGAAGTTGGGCGACAGCG 540
QY 607 TGCCTTCCCTGCGAAGGTTGTGTGATCTTCTGCGCACACAGGCTCATCTGTCCATCGT 666
DB 541 TGCCTTCCCTGCGAAGGTTGTGTGATCTTCTGCGCACACAGGCTCATCTGTCCATCGT 600
QY 667 GTGCTGATGATCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
DB 601 GTGCTGATGATCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 727 CTTGAGATATACCCAGGCAACAGAGCTTAACCTGCAGTACAGCTTGTGTAGTGTGG 786
DB 661 CTTGAGATATACCCAGGCAACAGAGCTTAACCTGCAGTACAGCTTGTGTAGTGTGG 720
QY 787 CCTCTCTCTGACGGAATGTCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 846
DB 721 CCTCTCTCTGACGGAATGTCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 847 CCGAACCCTGCTCCGCTTCCGCGGGGGCCATCTTAACATGCGCATTTAAAGATCTTTAA 906
DB 781 CCGAACCCTGCTCCGCTTCCGCGGGGGCCATCTTAACATGCGCATTTAAAGATCTTTAA 840
QY 907 GTTAAAGACATTTAAAGAGAAATCCCTGGGTGAGCTCATCAATTTGCTCCAAAGATG 966
DB 841 GTTAAAGACATTTAAAGAGAAATCCCTGGGTGAGCTCATCAATTTGCTCCAAAGATG 900
QY 967 GCAGAGAAATGTTGAGGAGAGCCGCTTGGAGCTGCTGCTGAGAGAGCCGCTTGTTC 1026
DB 901 GCAGAGAAATGTTGAGGAGAGCCGCTTGGAGCTGCTGCTGAGAGAGCCGCTTGTTC 960

QY 1027 CACCTTAGGCATGATTTATATATATATATATATATATATATATATATATATATAT 1086
DB 961 CACCTTAGGCATGATTTAT 1020
QY 1087 TGTATTTATCTCTCTTTTACCCAGCAATGATGTTGCTATCAGCGCTCCAGCATTTTCAG 1146
DB 1021 TGTATTTATCTCTCTTTTACCCAGCAATGATGTTGCTATCAGCGCTCCAGCATTTTCAG 1080
QY 1147 GAGAAAATCCGTGGCCCGCCAGGATGACGTGTCCAGAAATGAAATGAAATTTACTTA 1206
DB 1081 GAGAAAATCCGTGGCCCGCCAGGATGACGTGTCCAGAAATGAAATGAAATTTACTTA 1140
QY 1207 CATTAATTTATCAAAATATATCCCTGGGTCAAAAGCATTTTCTCAGAGTTCAGAAAT 1266
DB 1141 CATTAATTTATCAAAATATATATATATATATATATATATATATATATATATAT 1200
QY 1267 CCGCGAGAGAGAGCCGCGATATATGAAAAAGCCGGCTCTTCCAGAGCATCATCTGGG 1326
DB 1201 CCGCGAGAGAGAGCCGCGATATATGAAAAAGCCGGCTCTTCCAGAGCATCATCTGGG 1260
QY 1327 TGTGCTCCCATTTGCTGCTGATGATCCAGCGTGTGACCTTCTCTGCTCATATGACCT 1386
DB 1261 TGTGCTCCCATTTGCTGCTGATGATCCAGCGTGTGACCTTCTCTGCTCATATGACCT 1320
QY 1387 GGGCTTCATCTGACAGCAGCAGAGCTTTCACAGTGTGACAGCTTTCATTTCCATGAC 1446
DB 1321 GGGCTTCATCTGACAGCAGCAGAGCTTTCACAGTGTGACAGCTTTCATTTCCATGAC 1380
QY 1447 TTTTGTCTTGAAGTAAACCGTTTCACTAAAGTCCCTCCTCAGAGCCCTCAGTGTG 1506
DB 1381 TTTTGTCTTGAAGTAAACCGTTTCACTAAAGTCCCTCCTCAGAGCCCTCAGTGTG 1440
QY 1507 TGAAGATTTAAGATTTTCTTCTAATGGAAGGTTTCATGATTAAGAACAAACACAC 1566
DB 1441 TGAAGATTTAAGATTTTCTTCTAATGGAAGGTTTCATGATTAAGAACAAACACAC 1500
QY 1567 CAGTCTCAGATCAGATGAGATGAAAAATGCACCTTGGCATGGAGCTCTCCACATC 1626
DB 1501 CAGTCTCAGATCAGATGAGATGAAAAATGCACCTTGGCATGGAGCTCTCCACATC 1560
QY 1627 CAGTATCCAGAACTGCGCAAGCTGACCCCAAAATGAAAAAGAAAGAGGCTTCCAG 1686
DB 1561 CAGTATCCAGAACTGCGCAAGCTGACCCCAAAATGAAAAAGAAAGAGGCTTCCAG 1620
QY 1687 GGGCAAGAAAGAGAGGAGGAGGAGCTGACGCGACTGAGCATCAGGCGTGTGCGAGA 1746
DB 1621 GGGCAAGAAAGAGAGGAGGAGGAGCTGACGCGACTGAGCATCAGGCGTGTGCGAGA 1680
QY 1747 GCAGAAAGGCGCACCTCTCTGAGACGTACAGCGGCCCACTCCCGAAGAGAAAGAG 1806
DB 1681 GCAGAAAGGCGCACCTCTCTGAGACGTACAGCGGCCCACTCCCGAAGAGAAAGAG 1740
QY 1807 CAAGCATATCACCTGCGGCGCTTACAGAGAGACCTGACAGCATGATGATGGA 1866
DB 1741 CAAGCATATCACCTGCGGCGCTTACAGAGAGACCTGACAGCATGATGATGGA 1800
QY 1867 GATCCAAAGAGGTAACCTGTTGGAATCTGCGGCACTGTGGAAGTGAAGAAACCTCTCT 1926
DB 1801 GATCCAAAGAGGTAACCTGTTGGAATCTGCGGCACTGTGGAAGTGAAGAAACCTCTCT 1860
QY 1927 CATTTACAGCATTTTATGAGCAATGACGCTTCTAAGGGGACAGATGCAATGATGGAAC 1986
DB 1861 CATTTACAGCATTTTATGAGCAATGACGCTTCTAAGGGGACAGATGCAATGATGGAAC 1920
QY 1987 CTTCGCTTATGTGGCCACAGGCTGATTCCTCATATGCTACTTGTAGAGACAATCTCT 2046
DB 1921 CTTCGCTTATGTGGCCACAGGCTGATTCCTCATATGCTACTTGTAGAGACAATCTCT 1980
QY 2047 GTTGGGAAGAAATATGATGAAGAAAGATACAACTGTGTCTGAACAGCTGCTGCTGAG 2106
DB 1981 GTTGGGAAGAAATATGATGAAGAAAGATACAACTGTGTCTGAACAGCTGCTGCTGAG 2040

QY 2107 GCGTACCTGGCATCTCTCCACAGCAGACCTGACGAGATGGAAGCGACGACCAA 2166
|||||
Db 2041 GCGTACCTGGCATCTCTCCACAGCAGACCTGACGAGATGGAAGCGACGACCAA 2100
QY 2167 CCTGAGCGTGGGAGCGCCAGAGGATCAGCCTTGCCGGGCTTGTATGATGACAGAG 2226
|||||
Db 2101 CCTGAGCGTGGGAGCGCCAGAGGATCAGCCTTGCCGGGCTTGTATGATGACAGAG 2160
QY 2227 CATCTACCTCTGAGCAGACCCCTCAGTGCCTTAGANTGCCATGTGGCAACCATCTT 2286
|||||
Db 2161 CATCTACCTCTGAGCAGACCCCTCAGTGCCTTAGANTGCCATGTGGCAACCATCTT 2220
QY 2287 CAATAGTCTATCCGSGAAACATCTCAAGTCCAAAGACATCTGCTTTGTATACCAACAT 2346
|||||
Db 2221 CAATAGTCTATCCGSGAAACATCTCAAGTCCAAAGACATCTGCTTTGTATACCAACAT 2280
QY 2347 ACAGTACCTGTTGACTGTGATGAAGTATCTTCATGAAGAGGCTGTATTAAGGAAAG 2406
|||||
Db 2281 ACAGTACCTGTTGACTGTGATGAAGTATCTTCATGAAGAGGCTGTATTAAGGAAAG 2340
QY 2407 AGGCAACCATGAGGAACCTGATGAATTAATGATGATGATGATGATGATGATGAT 2466
|||||
Db 2341 AGGCAACCATGAGGAACCTGATGAATTAATGATGATGATGATGATGATGATGAT 2400
QY 2467 GTTCTGGGAGAGACACCGCAGTTGAGATGATGATGATGATGATGATGATGATGAT 2526
|||||
Db 2401 GTTCTGGGAGAGACACCGCAGTTGAGATGATGATGATGATGATGATGATGATGAT 2460
QY 2527 GAAGAAGTCAACAACAAGAGGTCTCTAAACAGATCAATTAAGAAGAAAGAGAGTAA 2586
|||||
Db 2461 GAAGAAGTCAACAACAAGAGGTCTCTAAACAGATCAATTAAGAAGAAAGAGAGTAA 2520
QY 2587 GCCAGAGAGAGGAGAGTGTGACAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2646
|||||
Db 2521 GCCAGAGAGAGGAGAGTGTGACAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAG 2580
QY 2647 AGTATATGCTGTATCAATCAACAGCAGCTGCTGGGGGCCCCCTTGGCATCTCTGATATATATG 2706
|||||
Db 2581 AGTATATGCTGTATCAATCAACAGCAGCTGCTGGGGGCCCCCTTGGCATCTCTGATATATG 2640
QY 2707 CCTTTTCATGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2766
|||||
Db 2641 CCTTTTCATGCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
QY 2767 CAAGCAAG 2826
|||||
Db 2701 CAAGCAAG 2760
QY 2827 CATGAAG 2886
|||||
Db 2761 CATGAAG 2820
QY 2887 CATGCTGATCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2946
|||||
Db 2821 CATGCTGATCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
QY 2947 CTCCGGGCTGATGAG 3006
|||||
Db 2881 CTCCGGGCTGATGAG 2940
QY 3007 CAGACACCCCAAG 3066
|||||
Db 2941 CAGACACCCCAAG 3000
QY 3067 GCGGCTGCGCTTCCAGAGCGGAGATGTTATCAGAGAGTATCCGAGTCTCTCTGCT 3126
|||||
Db 3001 GCGGCTGCGCTTCCAGAGCGGAGATGTTATCAGAGAGTATCCGAGTCTCTCTGCT 3060
QY 3127 GGGAGATGATGCGAGAGAGTCTTCCGAGTCTCTTGTGAGAGTGGGCGCCCTTGTATCT 3186
|||||
Db 3061 GGGAGATGATGCGAGAGAGTCTTCCGAGTCTCTTGTGAGAGTGGGCGCCCTTGTATCT 3120
QY 3187 CTTTTCAGTCTCTGACATGTTCTCCAGAGGCTCTGATTCGGAGAGCTGAAGCGCTGGAGCA 3246

Db 3121 CTTTTCAGTCTCTGACATGTTCTCCAGAGGCTCTGATTCGGAGAGCTGAAGCGCTTGGAGCA 3180
|||||
QY 3247 TATCAGCAGTACACCTTCTCTCTCCACATCAGCTCAGATACAGAGGCTTGGCACCAT 3306
|||||
Db 3181 TATCAGCAGTACACCTTCTCTCTCCACATCAGCTCAGATACAGAGGCTTGGCACCAT 3240
QY 3307 CCAGCCTTACAAATAAGGAGAGAGAGTCTCTGACAGATACAGAGAGCTGTGATGACAA 3366
|||||
Db 3241 CCAGCCTTACAAATAAGGAGAGAGAGTCTCTGACAGATACAGAGAGCTGTGATGACAA 3300
QY 3367 CCAAGCTCTTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 3426
|||||
Db 3301 CCAAGCTCTTTTCTTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 3360
QY 3427 CATCAGATCGCCCTCATACACACACAGGAGCTGATGATGATGATGATGATGATGAT 3486
|||||
Db 3361 CATCAGATCGCCCTCATACACACACAGGAGCTGATGATGATGATGATGATGATGAT 3420
QY 3487 TCCCCAAGCTATGCGGGCTCTGCGCATCTTATGCTGTCCAGTTAACGGGCTGTCCA 3546
|||||
Db 3421 TCCCCAAGCTATGCGGGCTCTGCGCATCTTATGCTGTCCAGTTAACGGGCTGTCCA 3480
QY 3547 GTTATGAGTCAAGTGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3606
|||||
Db 3481 GTTATGAGTCAAGTGGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
QY 3607 TCATCTATTAAGACTCTGCTCTTGGAGACACTGCGCAATTAAGAACAGAGCTCCCTC 3666
|||||
Db 3541 TCATCTATTAAGACTCTGCTCTTGGAGACACTGCGCAATTAAGAACAGAGCTCCCTC 3600
QY 3667 CCCTGACTGCCCCAGAGAGAGAGAGTACCTTTGAGAGAGAGAGAGAGAGAGAGAG 3726
|||||
Db 3601 CCCTGACTGCCCCAGAGAGAGAGAGTACCTTTGAGAGAGAGAGAGAGAGAGAGAG 3660
QY 3727 AAACCTCCCTCGCTCTTAAAGAAAGTATCTTCAAGATCAAGTAAAGAGAGATGG 3786
|||||
Db 3661 AAACCTCCCTCGCTCTTAAAGAAAGTATCTTCAAGATCAAGTAAAGAGAGATGG 3720
QY 3787 CATTTGGGGGAG 3846
|||||
Db 3721 CATTTGGGGGAG 3780
QY 3847 GAGTTATCTGAGAGCTCATCAAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3906
|||||
Db 3781 GAGTTATCTGAGAGCTCATCAAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
QY 3907 CGACCTCGAGAGCAAACTCTATCATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAG 3966
|||||
Db 3841 CGACCTCGAGAGCAAACTCTATCATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAG 3900
QY 3967 CAGATCAATTTGGACCCCTTCAACACAGTACAGTGAAGAGAGAGAGAGAGAGAGAG 4026
|||||
Db 3901 CAGATCAATTTGGACCCCTTCAACACAGTACAGTGAAGAGAGAGAGAGAGAGAGAG 3960
QY 4027 GAGGACACACATGAAGAGATGATGCTACCTCTGAACTGTGAATGAGATGAT 4086
|||||
Db 3961 GAGGACACACATGAAGAGATGATGCTACCTCTGAACTGTGAATGAGATGAT 4020
QY 4087 GGAAGATGGGATTAATCTCTAGTGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 4146
|||||
Db 4021 GGAAGATGGGATTAATCTCTAGTGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
QY 4147 GCTCGGACAGTGAAGATGATGATTTGATGAAGAGAGAGAGAGAGAGAGAGAGAG 4206
|||||
Db 4081 GCTCGGACAGTGAAGATGATGATTTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
QY 4207 AGACTTATGATTAAG 4266
|||||
Db 4141 AGACTTATGATTAAG 4200
QY 4267 TGCCATGCGCTGACACAGAGTCTAGAGCTCCGATAGAGATATGAGTGGGCGGAGAGCA 4326
|||||

| | | | |
|----|------|---|------|
| Db | 4201 | TGCGCAATGCGCTGCACACAGGTTCTAGCGTCGCATAGAGATTATAGTGTGCTGGCCACGGAGCA | 4260 |
| Qy | 4327 | GGTGTGGAGATTGACACCCCATCGGTCTCTTGTGCCACAGACAGTCCGCAATTCATATGC | 4386 |
| Db | 4261 | GGTGTGGAGATTGACACCCCATCGGTCTCTTGTGCCACAGACAGTCCGCAATTCATATGC | 4320 |
| Qy | 4387 | CATGTTGCTGTGCACAGAACAAAGTGGCTGTCAAGGCTGTACTTCCTCTTTGACGA | 4446 |
| Db | 4321 | CATGTTGCTGTGCACAGAACAAAGTGGCTGTCAAGGCTGTACTTCCTCTTTGACGA | 4380 |
| Qy | 4447 | AGTGTCTTTCTTTAGAGCATTCGCATTCGCCGCGCGGGGGGGCCCTTCATGCGCTGC | 4506 |
| Db | 4381 | AGTGTCTTTCTTTAGAGCATTCGCATTCGCCGCGCGGGGGGGCCCTTCATGCGCTGC | 4440 |
| Qy | 4507 | TCTACCGAATCCGTGCTTCGATTTTATCTTTTCGACAGCATTCGGATGGGCTT | 4566 |
| Db | 4441 | TCTACCGAATCCGTGCTTCGATTTTATCTTTTCGACAGCATTCGGATGGGCTT | 4500 |
| Qy | 4567 | GTTGTGTTCACTTTTAGGGAGATCATATTTTGATTTATTTATTTCCATATTCATGT | 4626 |
| Db | 4501 | GTTGTGTTCACTTTTAGGGAGATCATATTTTGATTTATTTATTTCCATATTCATGT | 4560 |
| Qy | 4627 | AAACAAATTTAGTTTGTGTTCTTAATTCACCTCTAAAGTTCCAGGAACCGTTATAT | 4686 |
| Db | 4561 | AAACAAATTTAGTTTGTGTTCTTAATTCACCTCTAAAGTTCCAGGAACCGTTATAT | 4620 |

```

P R I O R   A P P L I C A T I O N   D A T A :
A P P L I C A T I O N   N U M B E R :   08/141,893
F I L I N G   D A T E :   26-OCT-1993
C L A S S I F I C A T I O N :   435
P R I O R   A P P L I C A T I O N   D A T A :
A P P L I C A T I O N   N U M B E R :   08/407,207
F I L I N G   D A T E :   20-MAR-1995
C L A S S I F I C A T I O N :   435
A T T O R N E Y / A G E N T   I N F O R M A T I O N :
N A M E :   Steeg, Carol Miernicki
R E G I S T R A T I O N   N U M B E R :   39,539
R E F E R E N C E / D O C K E T   N U M B E R :   01546
T E L E C O M M U N I C A T I O N   I N F O R M A T I O N :
T E L E P H O N E :   (613) 545-2342
T E L E F A X :   (613) 545-6853
I N F O R M A T I O N   F O R   S E Q   I D   N O :   3 :
S E Q U E N C E   C H A R A C T E R I S T I C S :
L E N G T H :   5011 base pairs
T Y P E :   nucleic acid
S T R A N D E D N E S S :   double
T O P O L O G Y :   linear
M O L E C U L E   T Y P E :   CDNA
F E A T U R E :
N A M E / K E Y :   CDS
L O C A T I O N :   196..4788
US-08-463-092B-3

```

RESULT 4
 US-08-463-092B-3
 Sequence 3, Application US/08463092B
 Patent No. 576880
 GENERAL INFORMATION:
 APPLICANT: Cole, Susan P.C.
 APPLICANT: Deeley, Roger G.
 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
 TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PAREQ RESEARCH & DEVELOPMENT INNOVATIONS
 STREET: Queen's University at Kingston
 CITY: Kingston
 STATE: Ontario
 COUNTRY: CANADA
 ZIP: K7L 3N6
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,092B
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/966,923
 FILING DATE: 27-OCT-1992
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/029,340
 FILING DATE: 8-MAR-1993
 CLASSIFICATION: 435

| Query Match | 10.5% | Score 511.2 | DB 1 | Length 5011 |
|-----------------------|--|--------------------|-----------|-------------|
| Best Local Similarity | 52.5% | Pred. No. 1,4e-122 | | |
| Matches 1367 | Conservative 0 | Mismatches 1178 | Indels 59 | Gaps 9 |
| QY 1811 | CACATCCACCTGGGGCCACCTCGCTTTACAGAGAGACACTGCACACGATCATGATCTGGAGATC | 1870 | | |
| Db 2142 | CACATTCACCTGGGGCCA--GGAGCGACCCCTCCACACGAAATGCGATCATCCCTTCCCATC | 2199 | | |
| QY 1871 | CAAGGGGTAAACGCTTGGGAATCGGGCGGCGAGTGGGGGAAGGGAACCAACCTCTCATC | 1930 | | |
| Db 2200 | CCCGAAGGTGCTTGGTGTGGCCGTGGTGGGCCAGAGGGGGCTGGGGAATCTGCTCTGCTC | 2259 | | |
| QY 1931 | TCAGCCATTTTAAAGGCCAGCATAGCGTTCTTACAGAGGACACTTTCACATGAGGAACCTTC | 1990 | | |
| Db 2260 | TCAGGCCCTCTTGGCGTGAATGGAACAAATGGAGGGGGCAGCGTGCATATCAAGGGGCTCCGTC | 2319 | | |
| QY 1991 | GCTTATGTGGCCCCAGCAGCGCTGGATCTCTCAATGCTACTCTGAGAGAACATCTCTGTT | 2050 | | |
| Db 2320 | GCTTATGTGCCACACAGCAGCGCTGGATTTAGATGATTTCTTCCCGAAGAAACATCTTTT | 2379 | | |
| QY 2051 | GGGAAGAAATATGATGAGGAAGAAATACACACTGTGCTGCAAGCTGCTGCGCTGAGGCT | 2110 | | |
| Db 2380 | GGATGTCAAGCTGGAGGAACCATATTACAGGTTCGTGATACAGGCTGTGCTCTCTCCCA | 2439 | | |
| QY 2111 | GACCTGGCCATTTCTTCCAGCAGCAGCCTGACGAGATTTGGAGAGCAGAGGCCAACCTG | 2170 | | |
| Db 2440 | GACCTGGAATTCCTGGCCAGTGGGGATGCGAAGAGATTTGGCGAAGAGCGCGTGAACCTG | 2499 | | |
| QY 2171 | AGCGGTGGGCGAGCGCCAGAGATCAGCCTTGGCCCGGCGCTGTATAGTACAGAGAGATC | 2230 | | |
| Db 2500 | TCTGGGGGACGAAGACACGCGGTGACCTGGCGCGCGCTGTACTCCACGCGTCGACATT | 2559 | | |
| QY 2231 | TACATTCCTGGAGAGACCCCTCATGTCCTTTAATATGGCCATGGGCGCAACCATCTTTCAT | 2290 | | |
| Db 2560 | TACCTCTTCGATGATCCCCCTCTCAGCATGGATGGATGGCCATGTGGGAAACACATCTTTGAA | 2619 | | |
| QY 2291 | AGTGTAT-----CCGGAACATTCATCAATGCCAGACAGTTCGTGTTGTTACCCACAG | 2344 | | |
| Db 2620 | AATGTGATTTGGCCCCCAGAGGGGATGCTGAAGAAACAGACGCGGATCTTGTGTCACGCGACGC | 2679 | | |
| QY 2345 | TTACAGTTCCTGGTGTGATGTGAATGAATGATCTTCAATGAAGAAGAGGGGTGATTTACGGA | 2404 | | |
| Db 2680 | ATGAGCTACTTGGCCGAGTGGAGCTCATTCATCGTCATGATGATGGGGGCGCAAGATCTCTGAG | 2739 | | |
| QY 2405 | AGAGGCACCCATGAGGAACATGATGAATTTAAATGGTGAATATGCTACCATTTTAAATAC | 2464 | | |

Correspondence Address:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,109A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: PQI-002CP4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-462-109A-3

Query Match 10.5%; Score 511.2; DB 2; Length 5011;
Best Local Similarity 52.5%; Pred. No. 1,4e-122;
Matches 1367; Conservative 0; Mismatches 1178; Indels 59; Gaps 9;

1811 CACATCCACCTGGCCGACCTGGCTTACAGAGACATCGATCTGGAGATC 1870
1812 CACATCCACCTGGCCGACCTGGCTTACAGAGACATCGATCTGGAGATC 1870
2142 CACATCCACCTGGCCGACCTGGCTTACAGAGACATCGATCTGGAGATC 2199
1871 CAGAGAGGTAAGTGGTGAATCTGGGAGAGTGGGAAGTGAAGTGGTGAAT 1930
2200 CCGGAGAGTGGTGGTGGGAGAGTGGGAGAGTGGGAGAGTGGGAGAGTGG 2259
1931 TCAGCATTTCAGGCGCAGATGAGGCTTCTAGAGGCGCAGATTCAGTCACT 1990
2260 TCAGCCTTGGCTGAGTGAATGACAAAGTGAAGGCGCAGTGGCTATCAAG 2319
1991 GCTATGTGGCCGACGAGGCTGGATCTTCATATCTCTGAGAGACAATCTGTT 2050
2320 GCTATGTGGCCGACGAGGCTGGATTCAGAAATGATTCCTCGAAGAAAATCT 2379
2051 GGAAGAGATATGATGAGAGAAATGATCACTGTGCTGAAGCAGTGGCTGAG 2110
2380 GAGTGTGACCTGGAGAGACATATTCAGAGTCCGTATACAGGCTGTGCTCC 2439
2111 GACCTGGCATTCTTCCAGACGACCTGACGAGATTCGAGAGGAGGAGGCAACT 2170
2440 GACCTGGAATCTCTGCGCAGTGGGAGTGGACAGAGATTCGGGAGAAAGGCG 2499

2171 AGCGTGGGACAGGCGCAGAGATCAAGCTTGGCCGGGCTGTGATATGACAGGACATC 2230
2500 TCTGGGGACAGAAAGCAGCGCTGGAGCTGGCCGGGCGCTGACTCCACCTGACAT 2559
2231 TACATCTGGAGACCCCTCAGAGCTTAGATGCCATGTGGGCAACCACTTCAAT 2290
2560 TACCTTGTGATGATCCCTCTCAGAGTGAATGCCCATGTGGGAAACACATCTTTGA 2619
2291 AGTGTAT-----CCGAAACATCTCAAGTCCAAAGACAGTCTGTGTTTACCAC 2344
2620 AATGTGATGGCCCCAAGGGAGTGGTGAAGAAACAGACCGGATCTGTGTCAGCAC 2679
2345 TTAGAGTACCTGTTGACTGATGATGATGATGATGATGATGATGATGATGATG 2404
2680 ATGAGCTACTTCCCGCAGGTGAGCGTCAATGATGATGATGATGATGATGATG 2739
2405 AGAGCACCATGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 2464
2740 ATGGGCTCTACAGAGAGTGGTGGCTGAGAGCGGCGCTTGCTGATGCTGCTAC 2799
2465 CTGTGCTGGAGAGACACCCCACTT---GAGATCATTTCAAAAAAGAAACAGTGT 2521
2800 TATGCCAGCAGAGCAGAGCAGATGACAGAGAGAGAGAGAGAGAGAGAGAG 2859
2522 TCACAGAGAGATCACAGACAAG-----GTCTAAACAGGATCAATTAAGAG 2573
2860 CCAAGGAG 2919
2574 AAAAGCAGTAAAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2615
2920 CAAGTGCAGAGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2979
2616 AAGAGAAAGGAGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2675
2980 AACAGCAGCAG 3039
2676 GGGGG-----CCCTTGGCATTCTGTTTATGAGAGAGAGAGAGAGAGAGAG 2728
3040 GAGGCTGACAG 3099
2729 AGCAGCAGTTCAG 2788
3100 GCGATGGAATCTTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3159
2789 ACTGTGACT-----CGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2842
3160 GGGCTGGCTTCCAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3219
2843 CATATGAGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2902
3220 CAG 3279
2903 GCGATTCGAGG-----AGTGTCTTTCAGAGAGAGAGAGAGAGAGAGAGAG 2956
3280 GCGGTGTTGGCTACATCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3339
2957 CATGACAGACTTTCGAGAGAGATCTTCGAGAGAGAGAGAGAGAGAGAGAGAG 3016
3340 CAGTGTGAGCTGTGACAGAGATCTGCGGTACCCATGAGCTTCTTGAAGGAGAG 3399
3017 ACAGGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3076
3400 AGTGGAG 3459
3077 TTCAG 3136
3460 GAGGTATCAAGATGATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3519
3137 GAGAGAGCTTCCGAGAGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3196
3520 CTGCTGCGCAG 3579
3197 CTGACATGTCTCCAGAGAGTCTGATTCGAGAGAGAGAGAGAGAGAGAGAGAG 3256

DB 3580 GTCCAGAGGTTCTGCTGCTCTCCGCGACGCTGAAGCCCTGAGTGGGTAGCCGC 3639
QY 3257 TCACCTTTCTCTCCCATCATGCTCCAGCATATAGAGGCTTGCACATCCAGCCATC 3316
DB 3640 TCCCGGCTCTTCCCATTTAAGAGACCTTGGTGGGGGTACAGGTCATTCGAGCCCTC 3699
QY 3317 AATTAAGGCGAGAGTTTCTGACAGATACAGAGAGCTGTGATGACACCAAGCTCT 3376
DB 3700 GAGAGAGAGAGGCTCTATCATCAGAGAGTCACTGAAGGTGAGAGAGAGAGAGCC 3759
QY 3377 TTTTGTGTTAGGTGAGAGGAGGCTGAGCTGTGGGCTGAGGAGCTCATCATGATC 3436
DB 3760 TATTATCCACAGATCTGAGCAAGAGGCTGAGCGGCTGAGTGTGTGGGCAAC 3819
QY 3437 GCGCTCATCACACACAGGAGGCTGATGATCTTCTTATGACAGGAGATTCGCCACCC 3496
DB 3820 TGCATCTGTTGTTGCTGCTGCTGTTGCGGTGATCTCCAGGACAGGCTCATGCTGGC 3879
QY 3497 TATGCGGCTCTGCCATCTCTTATGCTGTCTCAAGTTAACGGGCTGTTCCAGTTACGGTC 3556
DB 3880 TTGGTGGGCTCTGAGTGTCTTACTCATTCAGAGTCAACAGTACTTGAACCTGGCTGTT 3939
QY 3557 AGACGTGATCTGAGAGAGAGTCACTCATCTGAGTGGAGAGATCAATCATCATCT 3616
DB 3940 CGGATGTATCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCATCA 3999
QY 3617 AAGACTCTGCTCTGAG 3676
DB 4000 GAGACTGAGAA--GAGAGGCGGCTGGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAG 4056
QY 3677 CCCAG 3736
DB 4057 CCCAG 4116
QY 3737 CTCGTCTCAAG 3796
DB 4117 TTCGTTCTCAGGAG 4176
QY 3797 CGGAG 3856
DB 4177 CGGAG 4236
QY 3857 GGAGGCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3916
DB 4237 GAG 4296
QY 3917 AGCAACTCTCTATCATCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3976
DB 4297 TTCAAGATCATCATCATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4356
QY 3977 TTGAG 4036
DB 4357 CTGAG 4416
QY 4037 ATGAAAG 4096
DB 4417 CTGAG 4476
QY 4097 GATTAAG 4156
DB 4477 GAG 4536
QY 4157 TGTAAAG 4216
DB 4537 ACGAAG 4596
QY 4217 ATTGAAG 4276
DB 4597 ATTCAG 4656
QY 4277 CTGAG 4336

DB 4657 CTCACACACATCATGAGTATACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4716
QY 4337 TTTGACACCCCATGAGTCTCTCTG 4360
DB 4717 TACGGCGCCCATGAG 4740

RESULT 6
US-08-460-907B-3
Sequence 3, Application US/08460907B
Patent No. 5891724
GENERAL INFORMATION:
APPLICANT: Deeley, Roger G.
APPLICANT: Cole, Susan P.C.
TITLE OF INVENTION: METHODS FOR CONFERRING MULTIDRUG
RESISTANCE ON A CELL
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: PARTED RESEARCH & DEVELOPMENT INNOVATIONS
STREET: Queen's University at Kingston
CITY: Kingston
STATE: Ontario
COUNTRY: CANADA
ZIP: K7L 3N6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,907B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: Q1551
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5011 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 196..4788
US-08-460-907B-3

Query Match 10.5%; Score 511.2; DB 2; Length 5011;
Best Local Similarity 52.5%; Pred. No. 1,4e-122;
Matches 1367; Conservative 0; Mismatches 1178; Indels 59; Gaps 9;

Db 4297 TTGAAGATCACCATCATCCCCAGAGACCTGTTTGTTCGGGTTCCTCCGATGAC 4356
Qy 3977 TTGAGCCCTTCAACAGTACACTGAAGACAGATTGGATGCCCTGGAGAGACAC 4036
Db 4357 CTGAGCCATTTACAGCCAGTCTGGATGAAGAGTGTGAGAGCTGAGCTGGCCAC 4416
Qy 4037 ATGAAGATGTATGTGTGACGCTACCTCTGAACTTGAATCTGAAGTGAAGATGG 4096
Db 4417 CTGAAGAGCTTCGTGTCAGCCCTTCGACAGTACAGTAAATGTGAAGAGGGGG 4476
Qy 4097 GATACCTTCAGTGGGGAGAGCCGCTTGTGATAGAGCCCTGCTCCGCCAC 4156
Db 4477 GAGACCTCAGTGTGCGGAGCCGCTTGTGCTAGCCCGGCTTGTGAGAG 4536
Qy 4157 TGTAGATCTGATTTAGATGAAGCCACAGCTCCATGACACAGAGACTTTTG 4216
Db 4537 ACGAAGATCTTGTGTTGATGAGGCGACGAGCCGTGAGCTGGAAGAGGAGCTC 4596
Qy 4217 ATTCAGAGACCATCCAGAGACTTTGACAGACTGTACATGCTGACCATTTGCCATCC 4276
Db 4597 ATCCAGTCCACATCCGAGACAGTTGAGAGACTGACACCGCTCCACCATCGCCACCG 4656
Qy 4277 CTGACAGGCTTCAGGCTCCGATAGATATGCTGCGCCGAGGAGAGGTGGAG 4336
Db 4657 CTACACACATCATGAGCTACACAGGCTGATCTTGGACAAAGAGATCCAGAG 4716
Qy 4337 TTGACACCCCATCGTCTCTCTG 4360
Db 4717 TACGGCGCCCATCGAGCTCTG 4740

RESULT
US-08-463-179A-3
; Sequence 3, Application US/08463179A
; Patent No. 6001563
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,179A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; APPLICATION NUMBER: 08/029,340
; FILING DATE: 8-MAR-1993
; APPLICATION NUMBER: 08/141,893
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: 08/407,207
; FILING DATE: 20-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConfil, Giulio A. Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: P01-0020P8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
; LENGTH: 5011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..4788
; US-08-463-179A-3

Query Match 10.5%; Score 511.2; DB 3; Length 5011;
Best Local Similarity 52.5%; Pred. No. 1.4e-122;
Matches 1367; Conservative 0; Mismatches 1178; Indels 59; Gaps 9;

Db 1811 CACATCCACCTGGGCGCCACCTGCTTACAGAGACACTGCACACATGATCTGAGATC 1870
Qy 2142 CACATTCACCTGGGCGCA--GGAGCGACCTCCACACTGAATGSCATCACCCTTCATC 2139
Db 1871 CAGAGGGTAAACTGTTGAATCTGCGGCGAGTGGGAAAGTGAAGAAACCTCTCATTT 1930
Qy 2200 CCGAAGGTGCTTGTGTGGCGCGGTGGCCAGGTGGGCTCGGAAAGTGTGCTGCTC 2259
Db 1931 TCAGCCATTTAGGCGCAGATGACCTTCTAGAGGCGACATTCGATTCAGTGAACCTTC 1990
Qy 2260 TCACCCCTTGTGGCTGAGATGACAAAGTGAAGGGGACGCTGCTATCAAGGCTCGTG 2319
Db 1991 GCTATGTGGCGCCAGAGGCGCTGATCTCTCAATGCTACTGTGAGAGCAACATCTGTT 2050
Qy 2320 GCTATGTGCGCCAGAGGCGCTGATCTCTCAATGCTACTGTGAGAGCAACATCTGTT 2379
Db 2051 GGAAGGAATATGATGAAGAAAGATACAACTGTGCTGAAGACGCTGCTGAGGCT 2110
Qy 2380 GGATGTGACGTGAGAGAACATATTAACAGTCCGTGATACAGGCTGCTGCTCTCCCA 2439
Db 2111 GACCTGGCCATTTCTCCACAGAGGACCTGAGAGATGAGAGCGAGCGCAACCTG 2170
Qy 2440 GACCTGAAATCTGCCACAGTGGGATGAGACAGATGAGCGGAGGAGGCTGACCTG 2499
Db 2171 AGCGTGGGCGAGCGCCAGAGAGATGACCTTCCCGGCGCTGTATGACAGGAGCATC 2230
Qy 2500 TCTGGGGAGACAGAGCGCGGTGAGCTGCGCGGCGGTGCTCCAGCGAGCATTT 2559
Db 2231 TACATCTGAGAGACCCCTCAGTGCCTTAGATGCCATGTGSCAACACCATCTTCAT 2290
Qy 2560 TACCTCTTGATGATCCCTCTCAGAGTGGATGCCATGTGGGAAACACATCTTGA 2619
Db 2291 AGTGTAT-----CGGAACATCTCAAGTCAAGTCAAGTCTGTGACCAACAG 2344
Qy 2620 AATGTGATGGCCCCAGAGGAGTCTGAAGACAGAGCGGATCTGTGACGACAGC 2679
Db 2345 TTACAGTACCTGTTGACTGTGATGAAGTATCTTCATGAAGAGGCTGTATTACGAA 2404
Qy 2680 ATGAGCTACTGGCGGAGAGTGAAGTCAATGATGATGAGTGGCGGAAGATCTGAG 2739
Db 2405 AGAGCACCCATGAGAGCTGATTAATTTAATGTGACTATGCTTACCATTTTAATAC 2464
Qy 2740 ATGGGCTCTTACAGAGGCTGCTGCTGAGACGCGCTTGTGAGTCTCTCGTAC 2799
Db 2465 CTGTGCTGGGAGAGACCCGCAAGT---GAGATCAATCAAAAAGGAAACAGTGT 2521
Qy 2800 TATCCAGACAGAGAGAGAGAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2859
Db 2522 TCACAGAGAGATCAACAGAGAG-----GTCTTAAACAGATCAATAAAGAG 2573
Qy 2860 CCAG 2919
Db 2574 AAAAG 2615
Qy 2920 CAATCGAG 2979
Db 2616 AAG 2675

| | | | |
|----|------|---|------|
| QY | 2291 | AGTCCTAT-----CCGGAAACATCTCAAGTCCAAAGACAGCTCTGTGTTGTAACCAACAG | 2344 |
| Db | 2620 | AATGTGATTGGCCCCAAGGGGAGTGTGAAGAAACAAAGCCGATCTTGGTCAAGCAGAC | 2679 |
| QY | 2345 | TTACAGTACCTGGTTGATCTGTATGAAGTATCTTCATCAAGAGGGCGTATTAACGAA | 2404 |
| Db | 2680 | ATGAGCTACTTGGCCGAGGTGGACGTATCATCTGTCTATGATGAGGGGCAAGATCTCTAG | 2739 |
| QY | 2405 | AGAGGCAACCCATGAGAACTGATGAATTTAAATGTTGACTATGCTACATTTTAAATAC | 2464 |
| Db | 2740 | ATGGGCTCTTACACAGGAGAGCTGTGGCTCGAGACGGGGCCTTGCTGATGAGTTCCTGGCTAC | 2799 |
| QY | 2465 | CTGTGGTGGGAGAGACCCGCAAGT---GAGATCAATTCAAAAAGAAACAGCTGT | 2521 |
| Db | 2800 | TATGCCACAGACAGAGCAGAGAGTGCAGAGAGAACGGGGTCAAGGGCGCTCAGCGT | 2859 |
| QY | 2522 | TCACAGAAAGATCTACAAAGACAAG-----GTCTTAAACAGAGATCAATAAAGAG | 2573 |
| Db | 2860 | CCAGAGAAAGAAAGAAAGCAATGAGAAATGGCATGCTGTGTACGACAGTGCAGAGAG | 2919 |
| QY | 2574 | AAAAAGCATAAAGCCAGAGGAAGGGCACTGTGTGAGCTGG----- | 2615 |
| Db | 2920 | CAATGCGAGAGACAGCTCACACACTCTCTCTCTAATAGTGGGAGCAATGACAGCACAC | 2979 |
| QY | 2616 | AAGGAAGGGCAGGGGTTCGTGCTCGGTGGAGTATATGCTGTCAATCCAGGCTGTG | 2675 |
| Db | 2980 | AACAGCACCCGACAACTGCGAAAGACTGAGCCACAGAGAGAGAGACCTGGAAAGCTGATG | 3039 |
| QY | 2676 | GGGGC-----CCCTTGGCATCTTGGTATATATGAGCCCTTTTCATGCTGATATAGGC | 2728 |
| Db | 3040 | GAGGCTGACAAAGGGCGAGACAGGGGAGTCAACCTTTCGTGTACTGGGACTACATGAG | 3099 |
| QY | 2729 | AGCACCGCCTTCAGCACCTGTGTGTGATGTTACTGATCAACAAAGAAAGCGGAAACC | 2788 |
| Db | 3100 | GCAATCGGACTCTTCATCTCTCCACAGATCTTCCTTTATGATGTAACCATGTGTC | 3159 |
| QY | 2789 | ACTGTGACT-----CGAGGAACGAGAGCCTCGGTGAGAGAGCATGAAGACAATCT | 2842 |
| Db | 3160 | GGCGTGGCTTCAACTATTGTGCTCAGCCTCTGTGACTGATGACCCATGCTCAACGGGACT | 3219 |
| QY | 2843 | CATATGCACTATATGCCACACTCTACGCCCTTCCATGGCAGTATGCTGATCTCTGAA | 2902 |
| Db | 3220 | CAGGAGCAACGAAGATCGGGTGAGAGGTCTATGAGAGCCCTGGGCAATTCAACAGGATC | 3279 |
| QY | 2903 | GCCATTCGAG-----AGTTGCTTTGTCAAGGGCAGCTCGAGCTCTCTCCGCTG | 2956 |
| Db | 3280 | GCCGTGTTTGGCTACTCTCATGAGGCCGTGTCCATCGGGGAGATCTTGCCCTCCGCTGTCTG | 3339 |
| QY | 2957 | CATAGCAGGCTTTCCGAGAGATCTTCGAAGGCCATGAAATGTTTATGACAGACCCC | 3016 |
| Db | 3340 | CACGTGACCTGCTGCACACACTCTCTGGGCTACCCATAGCTTCTTTGAGCGAGCCCC | 3399 |
| QY | 3017 | ACAGGAGAGATTCTCAACAGGTTTTCACAAAGATGATGAATGACGTGGCGTGGCG | 3076 |
| Db | 3400 | AGTGGAAACCTGGTGAACCCCTTCTCCAAAGAGCTGACACAGTGGACTCCATGATCCCG | 3459 |
| QY | 3077 | TTCCAGGCCGAGATGTTCAATCCAGAAAGTTATCTGCTGTCTCTGTGTGGAAATGATC | 3138 |
| Db | 3460 | GAGGTCAATCAAGATGTTCAATGAGGCTCCCTGTTCAAGCATGTGGTGCATCGATCTGTATC | 3519 |
| QY | 3137 | GCAGAGATCTTCCCGTGTCTCTGTGAGCAGTGGGGCCCTGTCTATCTCTTTCACTG | 3196 |
| Db | 3520 | CTGTGTGACCAAGCCCATCGCCGCAATCATCTCCGCCCTTGGGCTCATCTACTTCTTC | 3579 |
| QY | 3197 | CTGACAAATGCTCTCAGAGGTCTGATTTGGGAGCTGAACGCTGGGAATATCAAGGAG | 3266 |
| Db | 3580 | GTCACAGAGTTCTATAGTGGCTTCTCTCCCGCAGCTGAAAGGCGCTGAGTGGGTGAGCCGC | 3639 |
| QY | 3257 | TCACCTTTCTCTCCACATCAAGCTCAGATACAGAGGCTTTGCACCATCACCCCTAC | 3316 |
| Db | 3640 | TCCCGGCTATTTCCATTTCACAGAGACCTTGCTGGGGGTCAAGGCTATCTTGAGCCTTC | 3699 |
| QY | 3317 | AATAAAGGCGAGAGTTTCTGCACAGATACAGAGACTGCTGATGACAAACCAAGCTCT | 3376 |

| | | | |
|----|------|--|------|
| Db | 3700 | GAGAGCAGGAGCGCCTTCACTCCACCGAGAGTCACTGAAAGTGGAGACACCAAGAGGCC | 3759 |
| OY | 3377 | TTTTTTTGGTTAAGTGTGCGATGCGGTGGCTGTGCGCTGGACCTCATCAGATC | 3436 |
| Db | 3760 | TATTACCCCGAGCATCGTGGCCCAACAGGTGGCGCCGTGCGGCTGGAATGTGTGGGCAC | 3819 |
| OY | 3437 | GCCTCATCAACACACGSGGGGTGATGATCGTTCTTATCCAGGGCBAATTCCCGACCC | 3496 |
| Db | 3820 | TGCATCTCTGTGTTGGTGGCCCTGTTTGGGGGTATCTCCAGGCACGCTTAGTGTGGC | 3879 |
| OY | 3497 | TATCGGGGTCTCGGCATCTCTTATAGCTGTGCAGTAAACGGGGCTGTCCAGTTAACGTC | 3556 |
| Db | 3880 | TTGGTGGGCTCTCAGTGTGTTACTATTCAGAGTCACCACTACTTGAACGTGGCT | 3939 |
| OY | 3557 | AGACTGCGATCTGAGACAGAGCTCGATTCACTCGGTGGTGAGAGATCAATCACTATT | 3616 |
| Db | 3940 | CGGATGTCACTGTGAATGGAACCAACATATCGGGCCGTGGAGGCGCTCAAGAGATATCA | 3999 |
| OY | 3617 | AAGACTGTCTCTGGAAAGACCTGGCAATTAAGAAACAGCGCTCCCTCGATCGG | 3676 |
| Db | 4000 | GAGACTGAGAA--GGAGGGGCCCTGGCAATCCAGAGACACGTCGCCACGACGTGG | 4056 |
| OY | 3677 | CCCCAGAGGAGGAGAGTACCTTTGAGAAACGCGAGATGAGGTACCGAGAAACCTCCCT | 3736 |
| Db | 4057 | CCCCAGGTGGGCGAGTGGAAATTCCGGAATCTACTGCTCGGTACCGAGAGACCTGGAC | 4116 |
| OY | 3737 | CTCGTCTTAAAGAAAGTATCTTCCAGATCTAAACCTTAAAGACAAATTGGATTGTGGGG | 3796 |
| Db | 4117 | TTTCGTTCTCAGGGCCATCAATGTCAAGATCAATGAGGGGAAAGAGTGCGATCGTGGGG | 4176 |
| OY | 3797 | CGGACAGATATCAGGAGAGTCTCGTGGGGAGTGGGCCCTCTCCGTCGTGGTAGATTACT | 3856 |
| Db | 4177 | CGGACGGGAGCTGGGAAGTGTCTCTCGACCTTGACCTTGAGCTTATTCGGATCAAGATGTCC | 4236 |
| OY | 3857 | GGAGCTGCATCAAGATTGATGAGAGTGAAGATCCAGTATTTGGCTTGCAGCTCCGA | 3916 |
| Db | 4237 | GAGAGAGATCATCATCATGATGGCATCAACATCGCCAGATGCGGCTGCAGACCTCCG | 4296 |
| OY | 3917 | AGCAAACTCTCTATCATTTCTCTCAAGAGCGGTGCTTTAGTGGCAGTGTCAAGATCAAT | 3976 |
| Db | 4297 | TTTCAAGATTCACATCATCCCCCAGGAGCCGTGTTTGGTTCCTCGATGTAC | 4356 |
| OY | 3977 | TTGAGCCCTTCAACCGATCACTGGAAGCCAGATTGGGATGGCCCTGAGAGGACAC | 4036 |
| Db | 4357 | CTGAGCCCATTCACCCAGTACTCGTGGATGAAGAGTGTGGACCTCCCTGGAGCTGCCAC | 4416 |
| OY | 4037 | ATGAAAGATGATATGTCTAGCTACCTCTGAACCTTGAATCTGAAGTATGAGAGATGG | 4096 |
| Db | 4417 | CTGAAGGACTTTCGTGTACGCCCTTCTCGAACAGCTAGACATGAATGTGACAGAAAGCGGG | 4476 |
| OY | 4097 | GATPACTTCTCAGTGGGGGGAAGCGCAGCTTGTGCAATGCTAGAGCCCTGCTCCGCCAC | 4156 |
| Db | 4477 | GAGAACCTCAGTGTGCGGCGAGCCACAGCTTGTGTGCTGTGCCCGGGCCCTGCTGAGGAAG | 4536 |
| OY | 4157 | TGTGAAGATTCGATTTTAAGTGAAGCCACAGCGCCCATGGAACAGAGACAGCACTTATG | 4216 |
| Db | 4557 | ACGAAGATCTTGTGTGGATGAGGCCACGGGACGCCGTGGACTTGGAAACGGACACCTC | 4596 |
| OY | 4217 | ATTCAAAAGACATCCGAGAGAGATTTTGAGAGACTGTACCATCTGACCAATTGCCATCGC | 4276 |
| Db | 4597 | ATTCAGTTCACACATCCGGAACAGATTTCAGAGATGACCCGTCTCACCATGSCCACC | 4656 |
| OY | 4277 | CTGCAACGCTTCAAGGCTCCGATAGGATTTATGTGTGCGCCAGGACACAGGTGGTGGAG | 4336 |
| Db | 4657 | CTCAACACCATCATGAGCATCAACAGAGGTGATGTCTTGACAAAGAGAAATCCAGGAG | 4716 |
| OY | 4337 | TTTGACACCCCATCGGTCTTCTG | 4360 |
| Db | 4717 | TACGGCCCATCGACCTCTCG | 4740 |

RESULT 10

Db 3340 CACGTGACCTGCTGCACAGATCTCGGTCACACGACCTCTTTGAGCGGACCC 3399
OY 3017 ACAGGAGATTCATCAAGGTTTCCAAAGACATGATGAATGAGCTGGCTGGCG 3076
Db 3400 AGTGGAGCTGTGAACCGCTTCTCAAGAGCTGGACACAGTGGACCTCATATCCG 3459
OY 3077 TTCAGGCGGAGATGTTTCATCCAGAGATTAATCTGTGTTCTTCTGTGGGAATATC 3136
Db 3460 GAGGTATCAAGATGTTTCATGGGCTCGTTCAAAGCATTTGGTGGCTGCATGTTATC 3519
OY 3137 GCAGAGCTTCCCGTGTCTTGTGGCACTGGGCCCCCTTGATCTCTTTTCACT 3186
Db 3520 CTGTGGCCAGCGCATCGCCGATCATATCCGCCCTTGCCCTCATCTACTTCTTC 3579
OY 3197 CTGCAGATTTGCTCCAGAGGCTGATTTGGGAGCTGAAGCTGCAATATACAGAG 3256
Db 3580 GTCCAGAGTTTCTAGTGGCTCTCTCCGACAGCTGAAGGCCCTCGAGTGGTACGCC 3639
OY 3357 TCACCTTCTCTCCACATCCATCCAGATACAGGCTTGGCACCATCCAGCCTAC 3316
Db 3640 TCCCGGCTATTCATTCATCAAGACCTTGTGGGGGTCAAGCTCATTCGAGCCTTC 3699
OY 3317 AATTAAGGCGAGGTTTCTCCACAGATACAGAGCTGCTGGATGACACCAAGCTCT 3376
Db 3700 GAGAGGAGGAGCGCTTCATCCACAGATGACCTGAAGGTGACGAGAACCAAGGCC 3759
OY 3377 TTTTCTTTTGTATACGTCGATGCGTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 3436
Db 3760 TATTAACCCAGATCTGAGGCGCAAGAGTGGCTGGCTGGCTGGCTGGCTGGCTGG 3819
OY 3437 GCGCTCATCCACACGAGGCTGATGATGCTTATGACGAGGAGATATCCCAAGCC 3486
Db 3820 TGCATCGTTCTGTTGCTGGCTGTTGCGGTGATCTCAAGGACAGCCTCATGCTGG 3879
OY 3497 TATGCGGCTCGCGCATCTCTATGCTGTCCAGTTAAGGGGCTGTTCAATTTAGCTC 3556
Db 3880 TTGGTGGCTCTGATGCTCTACTATTCATGAGGTTCACACGATCTGAATGCTGGTT 3939
OY 3557 AGACTGGATCTGAGAGAGATGATCTGATCTGATCTGATCTGATCTGATCTGATCT 3616
Db 3940 CGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3999
OY 3617 AAGCTCTGCTTGGAGACACCTGCGCAATTAAGAACAGCTCCCTCCCTGATCTG 3676
Db 4000 GAGACTGAGAA--GAGAGCGCCCTGCGCAATTCAGAGAGACCTCCGCGAGAGCTGG 4056
OY 3677 CCCGAGGAGGAGAGAGTGAAGCTTGAAGCGAGAGATGAGTACCGAGAAACCTCCCT 3736
Db 4057 CCCGAGTGGGCGAGTGAATTCGGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4116
OY 3737 CTGCTCTTAAAGAAATATCTTCACGATCAAACTTAAGAGAAATTTGGATTTGGG 3796
Db 4117 TTGCTTTCAGGACATCAATATGTCAGATCAATGAGGAGAGAAAGTGGATGTTGG 4176
OY 3797 CGGACAGATGAGGAGTCTGCTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 3856
Db 4177 CGGACGAGAGTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4236
OY 3857 GGAGGCTGCATCAAGATTTGATGAGTGAATCAATGATATTTGGCTGCTGCTGCTGCT 3916
Db 4237 GAGGAGAGATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4296
OY 3917 AGCAAACTCTCTATCATCTCTCAAGAGCGGCTGCTTCAAGTGGCTGCTGCTGCTGCT 3976
Db 4297 TTCAAGATCAACATCAATCCCGAGACCTGTTTGTGTTGGGTTCCCTCGATGATAC 4356
OY 3977 TTGAGCCCTTCAACAGTCACTGAAGACCAAGATTTGGATGCTGAGAGAGACAC 4036
Db 4357 CTGAGCCATTCACAGTACTCTCGATGAAGAACTGAGAGCTGAGAGCTGAGCCAC 4416
OY 4037 ATGAAGATATATGCTCAAGCTACCTGAAACTTGAATGATGAGATGATGAGAAAGG 4096

Db 4417 CTGAGGACTCTGCTGACGCCCTTCTGACAGAGCTAGACATGATTTGCAAGAGCGG 4476
OY 4097 GATAACTTCTGAGGGGAGAGCGAGCTCTTGTGATAGTAGAGCCCTGCTCCGAC 4156
Db 4477 GAGAACTCACTGTCGGGAGAGCGGAGCTTGTGTGCTAGCGGGGCTGCTGAGAG 4536
OY 4157 TGTAAATCTGATTTTATGATGAAGCAAGCTGCTGATGAGACAGAGAGACTTATG 4216
Db 4537 ACAGAGATCTCTGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4596
OY 4217 ATTCAG 4276
Db 4597 ATTCAG 4556
OY 4277 CTGACAGAGGTTTATGAGCTCCGATGATGATTTGTTGCTGCTGCTGCTGCTGCT 4336
Db 4657 CTCAACACATCATGAGCTACAGAGGATGATGCTGTTGAGCAAGAGAGAAATCAAGAG 4716
OY 4337 TTGACACCCCATCGGCTCTGCTG 4360
Db 4717 TACGCGCCCATCGAGCTCTGCTG 4740

RESULT 11
US-08-462-109A-1
; Sequence 1, Application US/08462109A
; Patent No. 5882875
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P.C.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,109A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; APPLICATION NUMBER: 08/029,340
; FILING DATE: 8-MAR-1993
; APPLICATION NUMBER: 08/141,893
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: 08/407,207
; FILING DATE: 20-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Decont, Giulio A. Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: P01-002CP4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

| | | | |
|----|------|--|------|
| QY | 2405 | AGAGGACCCCATGGAGAACTGATGAAATTTAAATGGGATGATATGCTACCAATTTTAAATAC | 2464 |
| Db | 2740 | ATGGGCTCTCTACAGAGAGCTGCTGCTCGAGACGGCGCTTGCTGATGTTCTGCTGCC | 2799 |
| QY | 2465 | CTGTGTCTGGGAGAGACCCGCACTT---GAGATCAATTCAAAAAGGAAACAGTGGT | 2521 |
| Db | 2800 | TATGCCAGCAGACAGGCGAGGAGCGAGATGCGAGAGAAACGGGGGTACGGCGGTACGCGGT | 2859 |
| QY | 2522 | TCACAGAGAAGTTCACAGACAAG-----GTCTTAAACAGATATCAATAAAGAAG | 2573 |
| Db | 2860 | CCAGGAGGAGGAAACCAAGCAAAATGGAATGGAATGCGATGCTGTTACAGGACAGTGCAGGAAG | 2919 |
| QY | 2574 | AAAAAGCATGAAGCCAGAGGAAAGGGCAGCTTGTGAGCTGG----- | 2615 |
| Db | 2920 | CAATGTGAGAGACAGCTCAGACACACTCTCTCTCTATATGAGGGAGACATCAGCAGGCACAC | 2979 |
| QY | 2616 | AAGAGAAAGGGCAGGGGTCCTGCTGCTCGCTGAGTATATGTTGCTTACATCCAGGCTGCTG | 2675 |
| Db | 2980 | AMCAGCACCCGACAGACTGCGAAGAACTTGAGGCCAGAGAGAGAGAACTTGGAACTGTATG | 3039 |
| QY | 2676 | GGGG-----CCCTTGGCATTCGTGTTATATGAGCCCTTTTCATGCTGATATGAGC | 2728 |
| Db | 3040 | GAGGCTGACAAAGGCGGAGACAGGAGGCGATCAAGCTTTCGCTGACGGGACTACATGAAG | 3099 |
| QY | 2729 | AGCAGCGCTTTCAGCAGCTCTGTGTGTGAGTTACTGGATTAAGCAAGAAAGGCGGACACC | 2788 |
| Db | 3100 | GCCATTCGAGCTCTTCACTTCCTCTCCACAGATCTTCCTTTTCATGTATGAACTGCTCC | 3159 |
| QY | 2789 | ACGTGTACT-----GGAGGGAACGAGAGCTCGGTGATGAGCAGATGAAGCAATCT | 2842 |
| Db | 3160 | GGCTGTGCTTCCAACTATTTGGCTCAGCCTGTGACTGATGATGCCCATGCTCAAGGGGACT | 3219 |
| QY | 2843 | CATATGAGTACTATGCGACAGCATCTAGCGCCCTCTCCATGCGAGTATGCTGATCTCTGAA | 2902 |
| Db | 3220 | CAGAGCAGACAGAAAGTCGGCGGTGAGCGTATATGAGGCCCTGGGATTTTCAAAAGGATC | 3279 |
| QY | 2903 | GCCATTTGAG-----AGTTGCTTTTGTCAGAGGCGAGCTCGAGCTTCTCCGGCTG | 2956 |
| Db | 3280 | GCCGTGTTTGGCTACTTCAGTGGCCGTGTCATTCGGGGGATCTTGGCTCCCGCTGTCTC | 3339 |
| QY | 2957 | CATGACAGGCTTTCCGAAGATCTTCGAAGCCCTATGAATTTTATGACAGGACGCCC | 3016 |
| Db | 3340 | CACGTGACCTGCTGCACACATCTCGCGGTACCCATATAGCTTCTTGAAGCGACGCCC | 3399 |
| QY | 3017 | ACAAGGAGATTTCTCACAGAGTTTCCAAAGACATGATGAAGTTGACGTGGCGCTGCCG | 3076 |
| Db | 3400 | AGTGGGAACCTGGTGAACCGCTCTCCAAAGAGAGCTGACACAGTGGACCTCATGATCCG | 3459 |
| QY | 3077 | TTCCAGGCCAGATGTTCAATCCAGAAAGTTATCTGTGTTCTTCTGTGTGGAAATGATC | 3136 |
| Db | 3460 | GAGGTCAATCAAGATGTTCATGAGGCTCCCTGTTCATAGTATGTGTGCTGCATCGTATATC | 3519 |
| QY | 3137 | GCAGAGATCTTCCCGTGTCTCTTTGGGAGTGGGGGCCCTGTCACTCTTTTCACTC | 3196 |
| Db | 3520 | CTGCTGGCACCGCCATCGCCGCATATCATCCGCCCTCTGGCGCTCATCTTCTTTC | 3579 |
| QY | 3197 | CTGCAGATGTCTCCAGAGGCTCTGATTTGGGAGATGAACGTCGTGGACAAATATCAACGAG | 3256 |
| Db | 3580 | GTCCAGAGGTTTACAGTGGTCTCTCCCGGACGCTGAACGGCTCGAAGTGGGTACAGCCG | 3639 |
| QY | 3257 | TCACCTTTCTCTCCACATCAGCTCCAGATACAGAGGCGCTTTCACCATCCAGCTAC | 3316 |
| Db | 3640 | TCCCGGCTATTTCCATTTTCAACGAGACCTTGCTGGGGGTGAGGTCTTTGCAAGCTTTC | 3699 |
| QY | 3317 | AATTAAGGGCAGGATTTCTGACAGATATACAGAGAGCTGCTGGATGACAAACCAACTCTCT | 3376 |
| Db | 3700 | GAGAGCAGGAGCGCTTATCCACACAGAGTACCTGAAGGTGAGCAGAACCAAGAGGCC | 3759 |
| QY | 3377 | TTTTTTTGTTAAGTGTGAGTGCAGGTGGTGGCTGTGCTGGCTGGACCTCATACAGACTC | 3436 |
| Db | 3760 | TATTTACCCAGCAGTCTGGGCAACAGTGGGTCTGGCGCTGGAAGTGTGTGGGCAAC | 3819 |
| QY | 3437 | GCCCTCATCACACAGGGGCTGATGATGCTTTATGCACGGGCAATTTCCCGCAGCC | 3496 |

| | | | |
|----|------|---|-------|
| Db | 3820 | TGCATCGTTCTGTTTGGCTGCCCTTTGGGGTGATGTCGAGGCACAGCCTCAGTCTGGC | 387.9 |
| Oy | 3497 | TATCGGGGTCTGCCATCTCTTAAGCTGTCCAGTTAACGGGCTGTCCAGTTTACGGTC | 355.6 |
| Db | 3880 | TTGGTGGGGCCCTCAGTGTCTTACTCATTTGACAGTGTACCACTACTTGAACTGGCTGGTT | 393.9 |
| Oy | 3557 | AGACTGGCATCTGAGACAGAAAGCTGCATTCACCTCGTGGTGAAGAGTAAATACACTACTT | 361.6 |
| Db | 3940 | CGATGTGATCTGTAAGAAAGAACCAACATCTGGCCGTGTGAAGAGCTCAAGGATTTTCA | 399.9 |
| Oy | 3617 | AAGACTGTGTCCTTGGAAGACCTGCACAGATTAAACAAAGCTCTCCTCCCTACTG | 367.6 |
| Db | 4000 | GAGACTGAGAA - - - GAGAGGGCCCTGGCAAAATCCAGNAGACAGCTGCCGACGAGCTGG | 405.6 |
| Oy | 3677 | CCCCAGAGGAGAGGTGACCTTTGAAACGACAGAGTAGTACCGAGAAAACCTCCT | 373.8 |
| Db | 4057 | CCCCAGGTGGGCCGAGTGGAAATTCGGAAACTCTGCTCGCTACCGAGAGACCTGGAC | 411.8 |
| Oy | 3737 | CTCTGCTTAAAGAAAGTATCCTTCACAGTCAAAACCTAAAGAAAGTATGGACTTTGGG | 379.6 |
| Db | 4117 | TTTCGTCTCAGGGCATCAATATGTCAGATCAATGGGGGAAAGTAGTGCAGATCTGGGG | 417.6 |
| Oy | 3797 | CGCAGAGATCAGGAAGTCTCTGCTGGGGATGGCCCTTCCGCTGTGGTGGAGTTATCT | 385.6 |
| Db | 4177 | CGGAGGGAGCTGGGAAGTGTCTCCCTGACCCCTGGCTTATTTCCGATCAAGAGTCTCC | 423.6 |
| Oy | 3857 | GGAGGCTGCATCAAGATTGATGCAAGTACATCAGTATATTGGCTTGCAGACTCCGA | 391.6 |
| Db | 4237 | GAAAGAGAGATCATCATCGATGGCATCAATCGCCAAATGGGCTGCAGCACTCCGC | 429.6 |
| Oy | 3917 | AGCAACCTCTTATCATTCCTCAAGAGCCGGTCTGTAGTGCCACTGTAGATCAAT | 397.6 |
| Db | 4287 | TTCAAGATCACATCATTCCTCCCGAGACCTGTTTGTTTTGGGGTTCCTCCGAATGAC | 435.6 |
| Oy | 3977 | TTTGACCCCTTCAACAGTACTGGAAGACAGATTGGGATGCCCTGGAGAGACACAC | 403.6 |
| Db | 4357 | CTGGACCATTCACCCAGTACTCGGATGAAGAAGTGTGAGACCTCCCTGGAGCTGGCCAC | 441.8 |
| Oy | 4037 | ATGAAAGATGTATTGCTCAGCTTACCTCTGAAACTTGAAATCTGAAAGTATGGAATGG | 409.6 |
| Db | 4417 | CTGAAGGACTTCGTGAGCCCTTCTCGACAACTGACCAATGATGTCACAAAGCCGG | 447.6 |
| Oy | 4097 | GATPACTTCTCAGTGGGGGAAAGCGACACTTGTGATAGTACAGACCTGTCCGCCAC | 415.6 |
| Db | 4477 | GAGAACTCAGTGTGGGGGAGCCGACAGCTGTGTGCTGACCCGGCCCTGTGTGAGGAAG | 453.6 |
| Oy | 4157 | TGTAGATTCTGATTTTATGATGAAGCCACAGCTGCACATGAGACAGACAGACTTATTG | 421.8 |
| Db | 4537 | ACGAAATCTTGTGTGATGAGGACACGCGCAGCCGTGGACTGTGAACAGCAGCACTC | 459.6 |
| Oy | 4217 | ATTCAAGAGCACTCCGAGAAAGCATTTGGCACTGTACCATGTGACCTTGGCCATGGC | 427.6 |
| Db | 4587 | ATCCAGTCCACATCCCGGACACAGTTGAGAGATGACGCTCTCTACACTGCCCCACGG | 465.6 |
| Oy | 4277 | CTGCACAGCGTTTAAAGCTCCGATAGGATTTATGCTGGCCACGAGACAGGTGGTGAG | 433.6 |
| Db | 4657 | CTCAACACACATCATGAGCTACACAAAGGATGATGCTTGTGACAAAGGAAATTCACGAG | 471.8 |
| Oy | 4337 | TTTGACACCCCATGGTCTTTG | 436.0 |
| Db | 4717 | TACGGCGCCCATGGAGCTCTG | 474.0 |

RESULT 13 RESULT
US-08-463-179A-1
: Sequence 1, Application US/08463179A
: Patent No. 6001563
: GENERAL INFORMATION:
: APPLICANT: COLE, Susan P.C.
: APPLICANT: Dealey, Roger G.
: TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
: NUMBER OF SEQUENCES: 6


```

Db 3580 GTCCAGAGGTTCTAGTGGCTTCCTCCGGCAGCTGAAGCGCTGAGTCCGCTCAGCCGC 3639
QY 3257 TCACCTTCTCTCCACATCAGTCCAGCATACAGGCGCTTCCACCATCCAGCCCTAC 3316
Db 3640 TCCCGGCTCTATCCCATTTCAACAGACCTTGGCGGGGTACGCTCATTCGACCTTTC 3699
QY 3317 AATAAGGGCAGAGTTCTGACAGATACAGAGCGCTGCGATGACAAACCACTCTCT 3376
Db 3700 GAGGAGCAGAGAGCGCTTATCCACAGATGACCTGAAGGTGACGAGAACGAGAGGCC 3759
QY 3377 TTTTGTGTTTACGTGTCATGCGATGGGCTGGCTGGCTGGCTGACCTTCATCATT 3436
Db 3760 TATATCCCGACATCGTGGCCACAGGTGGCTGGCTGGCTGGCTGAGTGTGGGCAAC 3819
QY 3437 GCCCTCATCCACCAGCGGGGCTGATGATCGTTCTTATGACAGGGGAGATCCCGAGCC 3496
Db 3820 TGCACTCGTTCTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3879
QY 3437 TATGGGGGTCTGCGCATCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3556
Db 3880 TTGTGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3939
QY 3557 AGACTGCGATCGAGACAGAAAGCTGATTCACCTGCTGAGAGAGATCATCATAT 3616
Db 3940 CGGATGCTATCTGAATATGAAACCAACATCGTGGCGCTGGAGAGGCTCAGAGATTTCA 3999
QY 3617 AAGACTGTCTCTGGAAGACCTGCCAGAAATTAAGAACAAAGGCTCCCTCCCTGACTGG 3676
Db 4000 GAGACTGAGAA---GGAGGCGCGCTGGCAAAATCCAGAGACACGCGCCGAGAGCTGG 4056
QY 3677 CCCGAGGAGGAGAGTACCTTTGAGAACGAGAGATGAGTACCGAGAAACCTCCCT 3736
Db 4057 CCCCAGGAGGAGGAGTAAATCCGGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4116
QY 3737 CTCGCTCTAAAGAAATGTTCTTACGATCAAACTAAAGAGATTTGGCTTGGGG 3796
Db 4117 TTCCTCTCAGGACATCATATGTCACGATCAATGGGGGAGAAAGTGGCTGCTGCTGCTG 4176
QY 3797 CGGACGAGATCAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3856
Db 4177 CGGACGAGGAGTGGGAAATCGTCCCTGACCCCTGAGCTTATTTCCGATCAACGAGCTGGC 4236
QY 3857 GGAGGCTGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3916
Db 4237 GAAAGGAGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4296
QY 3917 AGCAAACTCTATCATCTCTCAAGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3976
Db 4297 TTCAAGATCACCATATCCCCAGAGCCCTGTTTGTGTTTGGGTTCCCTCGAATGAAAC 4356
QY 3977 TTGACCCCTTCAACAGTACACTGAAAGCAAGATTTGGGATGCCCTGGAGAGACACAC 4036
Db 4357 CTGAGCCATTCAGCCAGTACTCGATGAAAGATCTTGAAGCTCCCTGGAGCTGGGCCAC 4416
QY 4037 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4096
Db 4417 CTGAAGGACTTGTGTCAGCCCTTCTGCAAGCTAGACATGAATGTGCAAAAGCGCGG 4476
QY 4097 GATTAATCTCTCAGTGGGAGAGCGAGCTTGTGATCATGATGAGAGCCCTGCTCCGCCAC 4156
Db 4477 GAGAACTCAGTGTGGGAGCGGAGCGCTGTTGTGCTGAGCCCGGCGCTGCTGAGAGAG 4536
QY 4157 TGTAAATCTGATTTTAAAGTAAAGCAAGCTGCAATGAGACAGAGACAGACTTATTTG 4216
Db 4537 ACGAATCCTCTGTGTGATGATGAGCGAGCGAGCCGTGAGCTGGAACGAGACGCTC 4596
QY 4217 ATTCAAGAGACCATCCGAGAGACATTTGACAGCTGTACATGCTGACCTTGGCCATGCG 4276
Db 4597 ATCCAGTCCACCATCCGAGACAGATTTGAGAGACTGACGCTCCTCAACCTCCGACCGG 4656
QY 4277 CTGACAGGTTCTAGGCTCCGATAGATTAATGATGCTGCTGCTGCTGCTGCTGCTGCTG 4336

```

```

Db 4657 CTCACACCATCATGACTGATACACAGGATGATGCTTTGACAAAGAGAAATCCAGAG 4716
QY 4337 TTTGACACCCCATCGCTCTTCTG 4360
Db 4717 TACGGCGCCCATCGACCTCTCTG 4740

RESULT 14
US-08-461-384B-1
; Sequence 1, Application US/08461384B
; Patent No. 6025473
; GENERAL INFORMATION:
; APPLICANT: Cole, Susan P. C.
; APPLICANT: Dealey, Roger G.
; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
; STREET: Queen's University at Kingston
; CITY: Kingston
; STATE: Ontario
; COUNTRY: CANADA
; ZIP: K7L 3N6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,384B
; FILING DATE: 05-JUN-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,923
; FILING DATE: 27-OCT-1992
; APPLICATION NUMBER: 08/029,340
; FILING DATE: 8-MAR-1993
; APPLICATION NUMBER: 08/141,893
; FILING DATE: 26-OCT-1993
; APPLICATION NUMBER: 08/407,207
; FILING DATE: 20-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Steed, Carol Mternickl
; REGISTRATION NUMBER: 39,539
; REFERENCE/DOCKET NUMBER: 01547
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2342
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 196..4788
; US-08-461-384B-1

Query Match 10.4%; Score 506.4; DB 3: Length 5011;
Best Local Similarity 52.4%; Pred. No. 2.4e-121;
Matches 1364; Conservative 0; Mismatches 1181; Indels 59; Gaps 9;

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 9, 2001, 11:40:32 ; Search time 3439.85 Seconds
(without alignments)
13319.761 Million cell updates/sec

Title: US-09-528-031-1

Perfect score: 4847
Sequence: 1 GCGTCATGCTCGCGGAGCGTG.....AAAAAAAAAGGCGGCGCCG 4847

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_hlc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 797.2 | 16.4 | 862 | 108 | AUI32064 AUI32064 |
| 2 | 686.6 | 14.2 | 761 | 108 | AUI32064 AUI32064 |
| 3 | 677.8 | 14.0 | 889 | 105 | AL543173 AUI32064 |
| 4 | 662.8 | 13.7 | 960 | 105 | AL543173 AUI32064 |
| 5 | 628.6 | 13.0 | 1206 | 189 | U66687 HSD6687 |
| 6 | 570.4 | 11.8 | 573 | 1672 | BE386272 BE386272 |
| 7 | 569.2 | 11.7 | 573 | 1672 | BE386272 BE386272 |
| 8 | 564.6 | 11.6 | 607 | 169 | BE386272 BE386272 |
| 9 | 542.4 | 11.2 | 545 | 166 | BE386272 BE386272 |
| 10 | 511.4 | 10.6 | 801 | 104 | AUI32064 AUI32064 |
| 11 | 478.4 | 9.9 | 495 | 15 | AI023781 AUI32064 |
| 12 | 465.6 | 9.6 | 520 | 139 | BE754240 BE754240 |
| 13 | 444.4 | 9.4 | 559 | 139 | BE754240 BE754240 |
| 14 | 444.2 | 9.2 | 529 | 119 | AM655995 AM655995 |
| 15 | 434.6 | 9.0 | 584 | 120 | AM761864 AM761864 |
| 16 | 430.4 | 8.9 | 451 | 17 | AI248060 AI248060 |
| 17 | 429.8 | 8.9 | 447 | 10 | AA649161 AA649161 |
| 18 | 429.4 | 8.9 | 511 | 139 | BE749194 BE749194 |
| 19 | 423.2 | 8.7 | 439 | 157 | HI7207 HI7207 |
| 20 | 418.4 | 8.6 | 420 | 16 | AI141646 AI141646 |
| 21 | 410.6 | 8.5 | 432 | 104 | BE698869 BE698869 |
| 22 | 408.8 | 8.4 | 594 | 104 | AUI32064 AUI32064 |
| 23 | 408 | 8.4 | 431 | 10 | AA704677 AA704677 |
| 24 | 408 | 8.4 | 496 | 9 | AA570953 AA570953 |
| 25 | 402.8 | 8.3 | 449 | 21 | AI493704 AI493704 |
| 26 | 397.4 | 8.2 | 450 | 158 | H60893 H60893 |
| 27 | 396.4 | 8.2 | 608 | 16 | AI132311 AI132311 |
| 28 | 386.4 | 8.0 | 807 | 175 | BE261806 BE261806 |
| 29 | 384.4 | 8.0 | 486 | 116 | AM480230 AM480230 |
| 30 | 385.6 | 8.0 | 1011 | 172 | BE024188 BE024188 |
| 31 | 383.6 | 7.9 | 431 | 17 | AI242936 AI242936 |
| 32 | 383.4 | 7.9 | 415 | 115 | AM379410 AM379410 |
| 33 | 382.2 | 7.9 | 478 | 151 | BE600246 BE600246 |
| 34 | 380.8 | 7.9 | 403 | 140 | BE812778 BE812778 |
| 35 | 377.8 | 7.8 | 410 | 12 | AA829904 AA829904 |
| 36 | 375.4 | 7.7 | 523 | 2 | AA124602 AA124602 |
| 37 | 371.6 | 7.7 | 428 | 149 | BF460507 BF460507 |
| 38 | 369.6 | 7.6 | 409 | 15 | AI032708 AI032708 |
| 39 | 360.8 | 7.4 | 566 | 14 | AA986339 AA986339 |
| 40 | 358 | 7.4 | 551 | 13 | AA930483 AA930483 |
| 41 | 356.4 | 7.4 | 385 | 159 | N46222 N46222 |
| 42 | 356 | 7.3 | 373 | 170 | BF866575 BF866575 |
| 43 | 346.8 | 7.2 | 350 | 171 | BF956728 BF956728 |
| 44 | 346.4 | 7.1 | 413 | 141 | BE860035 BE860035 |
| 45 | 343.8 | 7.1 | 483 | 13 | AA874464 AA874464 |

ALIGNMENTS

| RESULT | LOCUS | DEFINITION | ACCESSION | VERSION | KEYWORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS | TITLE |
|--------|----------|---|-----------|------------|-------------|--------|---------------------|--------------------|---|------------------------|
| 1 | AUI32064 | AUI32064 NT2RP3 Homo sapiens cDNA clone NT2RP3003721 5', mRNA sequence. | AUI32064 | AUI32064.1 | GI:10992418 | EST. | Homo sapiens human. | 1 (bases 1 to 862) | Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S., and Isogai, T. | HRI human cDNA project |

JOURNAL COMMENT

Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3951
 Fax: 81-438-52-3952
 Email: genomics@hri.co.jp
 HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

location/Qualifiers
 1..862
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NT2RP3003721"
 /clone_lib="NT2RP3"
 /cell_type="teratocarcinoma"
 /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
 BASE COUNT 188 a 230 c 227 g 213 t 4 others
 ORIGIN

Query Match 16.4%; Score 797.2; DB 108; Length 862;
 Best Local Similarity 95.1%; Pred. No. 1.3e-189;
 Matches 820; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| QY | 3008 | ACGACCCCGACGAGGAGGAGTTCCTCAACAGCTTTCCAAAGACATGATGATGACGCG | 3067 |
| DB | 1 | ACGACCCCGACGAGGAGGAGTTCCTCAACAGCTTTCCAAAGACATGATGATGACGCG | 60 |
| QY | 3068 | CGGCTCGGCTCCAGCGAGATGTTTCATCCAGAAAGTTATCTGTTCTGTGTG | 3127 |
| DB | 61 | CGGCTCGGCTCCAGCGAGATGTTTCATCCAGAAAGTTATCTGTTCTGTGTG | 120 |
| QY | 3128 | GGAATGATCGAGAGTCTCCCGGTGTTCTTGGGAGTGGGGCCCTTGCATCTTC | 3187 |
| DB | 121 | GGAATGATCGAGAGTCTCCCGGTGTTCTTGGGAGTGGGGCCCTTGCATCTTC | 180 |
| QY | 3188 | TTTTCAGTCTCCATTTGTTCTCCAGAGGCTCTGATTCGGAGCTGAGCGTGCAGAT | 3247 |
| DB | 181 | TTTTCAGTCTCCATTTGTTCTCCAGAGGCTCTGATTCGGAGCTGAGCGTGCAGAT | 240 |
| QY | 3248 | ATTCAGCAGTCACTTCTCTCCACATCAGCTCCAGATACAGAGGCTTCCACCATC | 3307 |
| DB | 241 | ATTCAGCAGTCACTTCTCTCCACATCAGCTCCAGATACAGAGGCTTCCACCATC | 300 |
| QY | 3308 | CACGCTCAATTAAGGCGAGAGTTTCTGCACATATACAGAGCTGCTGATGATCAGC | 3367 |
| DB | 301 | CACGCTCAATTAAGGCGAGAGTTTCTGCACATATACAGAGCTGCTGATGATCAGC | 360 |
| QY | 3368 | CAAGCTCTTTTCTTTTCTTACGTGAGATGCGGTGCTGCGGTGAGACCTC | 3427 |
| DB | 361 | CAAGCTCTTTTCTTTTCTTACGTGAGATGCGGTGCTGCGGTGAGACCTC | 420 |
| QY | 3428 | ATTCAGCAGTCCCTCATCCACACAGAGGCTGATGATGTTCTTATGACAGGCGAT | 3487 |
| DB | 421 | ATTCAGCAGTCCCTCATCCACACAGAGGCTGATGATGTTCTTATGACAGGCGAT | 480 |
| QY | 3488 | CCCCCAGCTATGCGGCTTCCCATCTTATGCTGTCCAGTTAACGCGGCTGTCCAG | 3547 |
| DB | 481 | CCCCCAGCTATGCGGCTTCCCATCTTATGCTGTCCAGTTAACGCGGCTGTCCAG | 540 |
| QY | 3548 | TTTACGTCAGCTGCGCATCTGACAGAGAGCTGATTCACCTCGGTGAGAGATCAAT | 3607 |
| DB | 541 | TTTACGTCAGCTGCGCATCTGACAGAGAGCTGATTCACCTCGGTGAGAGATCAAT | 600 |
| QY | 3608 | CACATCAATTAAGCTCTGCTTGGAGACCTGCGAGATTAAAGAACAGGCTCCCTCC | 3667 |
| DB | 601 | CACATCAATTAAGCTCTGCTTGGAGACCTGCGAGATTAAAGAACAGGCTCCCTCC | 660 |

| QY | 3668 | CCTGACTGGCCCCGAGAGAGAGAGTGCATCTTGGAGACCGAGATAGGATACCGAGAA | 3727 |
|---------------------------|--|--|---------------------|
| Db | 661 | CCTGACTGGCCCCGAGAGAGAGAGTGCATCTTGGAGACCGAGATAGGATACCGAGAA | 720 |
| QY | 3728 | AACTCCCTCTGCTGCTTAAAGAAAGTATCTTCACGATCAAACTTAAAGAGAAGTTGGC | 3787 |
| Db | 721 | AAACCTTCCTCTGCTGCTTAAAGAAAGTATCTTCACGATCAAACTTAAAGAGAAGTTGGC | 780 |
| QY | 3788 | ATTGTGGGCGGGAAGAGATACAGGGAAGTCGCGTGGGGATAGCCCCCTTCCTGTTGGT | 3847 |
| Db | 781 | ATTGTGGGCGGGAAGAGATACAGGGAAGTCGCGTGGGGATAGCCCCCTTCCTGTTGGT | 840 |
| QY | 3848 | GAGTTATCTGGAGGCGTCATCA | 3869 |
| Db | 841 | GTTATCTTGGAGGCGTCATCA | 862 |
| RESULT | 2 | | |
| LOCUS | AUI37090 | | |
| DEFINITION | AUI37090 | 761 bp | mRNA |
| ACCESSION | AUI37090 | | EST |
| VERSION | AUI37090.1 | | 24-OCT-2000 |
| KEYWORDS | sequence. | | |
| SOURCE | AUI37090.1 | GI:10997629 | |
| ORGANISM | human. | | |
| REFERENCE | 1 | | |
| AUTHORS | Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isogai, T. | | |
| TITLE | HRI human cDNA project | | |
| JOURNAL | Unpublished (2000) | | |
| COMMENT | Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. | | |
| FEATURES | | | |
| source | Location/Qualifiers | | |
| | 1..761 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone="PLACE1005753" | | |
| | /clone_lib="PLACE1" | | |
| | /tissue_type="placenta" | | |
| | /note="Vector: pHE15F13" | | |
| BASE COUNT | 200 a | 185 c | 204 g |
| ORIGIN | | 165 t | 7 others |
| Query Match | 14.2% | Score 686.6; | DB 108; Length 761; |
| Best Local Similarity | 96.4% | Pred. No. 8.8e-162; | |
| Matches 721; Conservative | 0; | Mismatches 24; | Indels 3; Gaps 2; |
| QY | 3650 | AAGAACAAGGCTCCCTCCCTGACGAGGCCCGCAGAGAGAGAGTACCTTTGAGAAGCA | 3709 |
| Db | 1 | AAGAACAAGGCTCCCTCCCTGACGAGGCCCGCAGAGAGAGAGTACCTTTGAGAAGCA | 60 |
| QY | 3710 | GAGATGAGGTACCGAAGAAACCTCCCTCTGCTCTTAAAGAAAGTATCTTCACGATCAA | 3769 |
| Db | 61 | GAGATGAGGTACCGAAGAAACCTCCCTCTGCTCTTAAAGAAAGTATCTTCACGATCAA | 120 |
| QY | -3770 | CCTAAAGAGAAGATTGGATTTGGGGGCGACAGATACAGGAATCTCGTGGGGATG | 3829 |

| | | | |
|------------|---|--|------|
| Dn | 121 | CCTAAGAGAAGATTGGCATTTGTG6GGGGGACAGAGATCAAGGAAGTCCCTCGTG6GGATG | 180 |
| OY | 3830 | GCCCTTCCTCCTCTGTGTGTGAGATTATCTGTGAGGCCTGCATCAAAGATTGATGGAGTAGAAATC | 3889 |
| Dn | 181 | CCCCCTCCCTCTGTGTGTGAGATTATCTGTGAGGCCTGCATCAAAGATTGATGGAGTAGAAATC | 240 |
| OY | 3890 | AGTAAATATGGCTCTGTGCCGACCTCCGAAAGCAAACTCTCTATCATCTCCCAAGAGCCGGTG | 3949 |
| Dn | 241 | AGTAATATGTGGCTGTGTGCCGACCTCCGAAAGCAAACTCTCTATCATCTCCCAAGAGCCGGTG | 300 |
| OY | 3950 | CTGTTCAGTGTGCACCTGTGCATCAAAATTTGGACCCTTCAACCAAGTACACTGAAGACCAG | 4009 |
| Dn | 301 | CTGTTCAGTGTGCACCTGTGCATCAAAATTTGGACCCTTCAACCAAGTACACTGAAGACCAG | 360 |
| OY | 4010 | AATTTGGAGTGGCTCTGTGAGAGACACACATGAAGAATGATTGTCTCACTACTCTGTAAA | 4069 |
| Dn | 361 | AATTTGGAGTGGCTCTGTGAGAGAGACACACATGAAGAATGATTGTCTCACTACTCTGTAAA | 420 |
| OY | 4070 | CTTGAATCTGAAAGTATGAGTGAATAATGGGGATTAATCTTCAGTGGGGGAACGGCACGCTCTG | 4129 |
| Dn | 421 | CTTGAATCTGAAAGTATGAGTGAATAATGGGGATTAATCTTCAGTGGGGGAACGGCACGCTCTG | 480 |
| OY | 4130 | TGCATAGCTTAAGGCCTGTGCTCGGCACACTGTAAAGATTCTGATTTAGATGAAGCCACAGCT | 4189 |
| Dn | 481 | TGCATAGCTTAAGGCCTGTGCTCGGCACACTGTAAAGATTCTGATTTAGATGAAGCCACAGCT | 540 |
| OY | 4190 | GCCATGACACACAGACAGACTTATGATTCAAGAGACATCCGAGAACGATTGGACAGC | 4249 |
| Dn | 541 | GCCATGACACACAGACAGACTTATGATTCAAGAGACATCCGAGAACGATTGGACAGC | 600 |
| OY | 4250 | TGTACACATGTCGACCATGGCCATGCGCTGACACAGGTTCAAGCTCCGANA-6GATTAT | 4308 |
| Dn | 601 | TGTACACATGTCGACCATGGCCATGCGCTGACACAGGTTCAAGCTCCGANAAGGATTAT | 660 |
| OY | 4309 | GGTCTGTGGCCGAGGAGACAGTGTGTGAGATTGTGACACCCCATCG--TCCTCTGTCCAAC | 4366 |
| Dn | 661 | GGGGCTGGGGCCCAAGACAGAGTGTGTGAGATTGTGNCACACCCATCGGGTCTCTCCNNGCCAAC | 720 |
| OY | 4367 | GACAGTTCCGAGTTCTGTATGCCATGTTTTG | 4394 |
| Dn | 721 | GACAGTTCCGAGTTCCNATGCCAAGTTG | 748 |
| RESULT | 3 | | |
| LOCUS | AL543173 | | |
| DEFINITION | AL543173 LTR NF1006_PL2 Homo sapiens cDNA clone CS0D1002YC13 5 prime, mRNA sequence. | | |
| ACCESSION | AL543173 | | |
| VERSION | AL543173.1 | GI:12875651 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo. | | |
| TITLE | 1 (bases 1 to 889) | | |
| JOURNAL | L1,W.B., Gruber,C.; Jesse,J. and Polayes,D. | | |
| COMMENT | Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr. Location/Qualifiers 1..889 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0D1002YC13" /clone_1kb="LTR_NF1006_PL2" /tissue_type="placenta" /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and | | |

Cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : filiang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 208 a 218 c 234 g 208 t 1 others

ORIGIN

Query Match 14.0%: Score 677.8; DB 106; Length 889;

Best Local Similarity 99.0%: Pred. No. 1.5e-159; Mismatches 7; Indels 0; Gaps 0;

Matches 682; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 28 GCGTGGCCGGGTGTCTCTGAGACAGGGGCGCAGAAATCTGATGTGAACCTAACAGTCTG 87
 Db 1 GCGTGGCCGGGTGTCTCTGAGACAGGGGCGCAGAAATCTGATGTGAACCTAACAGTCTG 60
 QY 88 TGAGCCCTTGGAACTCCACTCAGAGAGATGAGATATTCACATAGAAAAGATATAT 147
 Db 61 TGAGCCCTTGGAACTCCACTCAGAGAGATGAGATATTCACATAGAAAAGATATAT 120
 QY 148 CATCCCCAGTCTGGGTATAGAGTGTGAGGAGAGAACACAGCACTTCTGGAGCGCACAG 207
 Db 121 CATCCCCAGTCTGGGTATAGAGTGTGAGGAGAGAACACAGCACTTCTGGAGCGCACAG 180
 QY 208 AGACCGTGAAGATTCACAGATTCAGAGAACTGCACCTTGGATGCCAAGATGCTTGA 267
 Db 181 AGACCGTGAAGATTCACAGATTCAGAGAACTGCACCTTGGATGCCAAGATGCTTGA 240
 QY 268 AACAGACCCGAGCGGAGGCGCTCTCTGTGATGCTTCATGCATTCCTCAGCTAGAAAT 327
 Db 241 AACAGACCCGAGCGGAGGCGCTCTCTGTGATGCTTCATGCATTCCTCAGCTAGAAAT 300
 QY 328 CCTGGATGAGAGCATCCCAAGGAGAAATACCATCATGCTTGAATGCTGAGAGCCCAT 387
 Db 301 CCTGGATGAGAGCATCCCAAGGAGAAATACCATCATGCTTGAATGCTGAGAGCCCAT 360
 QY 388 CCGGACTACTTGCACAAACACAGACCCAGTGAACATCTGGGCTTTTCTCTGTATGAC 447
 Db 361 CCGGACTACTTGCACAAACACAGACCCAGTGAACATCTGGGCTTTTCTCTGTATGAC 420
 QY 448 TTTTTCGAGGCTTCTCTCTGAGGCGGCTGAGGCGGCGCACAAGAGGGGAGCTCTCAATGA 507
 Db 421 TTTTTCGAGGCTTCTCTCTGAGGCGGCTGAGGCGGCGCACAAGAGGGGAGCTCTCAATGA 480
 QY 508 AGAGCTGTGTCTCTCTCTGACAGCAGAGTCTTCTGACGTGAACCTCAGAAAGACTAGAG 567
 Db 481 AGAGCTGTGTCTCTCTCTGACAGCAGAGTCTTCTGACGTGAACCTCAGAAAGACTAGAG 540
 QY 568 ACTGTGCAAGAGAGCTGAATGAATGAGTGGGCGCAGAGCTGCTTCCGCGAAGGGTGT 627
 Db 541 ACTGTGCAAGAGAGCTGAATGAATGAGTGGGCGCAGAGCTGCTTCCGCGAAGGGTGT 600
 QY 628 GTGATCTTCTGCGGCGCAGAGCTCATCTGTCATCTGTCGCTGAGATGATACCCAGCT 687
 Db 601 GTGATCTTCTGCGGCGCAGAGCTCATCTGTCATCTGTCGCTGAGATGATACCCAGCT 660
 QY 688 GCGTGGCTTCAAGTGAAGCAGGCTTCATCG 716
 Db 661 GCGTGGCTTCAAGTGAAGCAGGCTTCATCG 689

RESULT 4

AL524917

LOCUS AL524917 960 bp mRNA EST 13-FEB-2001

DEFINITION AL524917 LTR.NFL003.NBC3 Homo sapiens cDNA clone CS00C009Y621 5

ACCESSION AL524917

VERSION AL524917.1 GI:12788410

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 960)

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 960

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS00C009Y621"

/clone_lib="LTR.NFL003.NBC3"

/sex="male"

/tissue_type="neuroblastoma cells"

/lab_host="DH10B"

/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dt) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed

by Life Technologies. Contact : Feng Liang Life

Technologies, a division of Invitrogen 9800 Medical Center

Drive Rockville, Maryland 20850, USA Fax : (1) 301 610

8371 Email : filiang@lifetech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 234 a 231 c 270 g 224 t 1 others

ORIGIN

Query Match

Best Local Similarity 98.8%: Pred. No. 9.2e-156;

Matches 678; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 31 TGGCGGGTGTCTCTGAGACAGGGGCGCAGAAATCTGATGTGAACCTAACAGTCTGCA 90
 Db 1 TGGCGGGTGTCTCTGAGACAGGGGCGCAGAAATCTGATGTGAACCTAACAGTCTGCA 59
 QY 91 GCGGCGGAGACCTCCACTCAGAGAGATGAGATATTCGATAGAGAAAAGATATATCAT 150
 Db 60 GCGGCGGAGACCTCCACTCAGAGAGATGAGATATTCGATAGAGAAAAGATATATCAT 119
 QY 151 CCCAGTCTCTGGGTATAGAGTGTGAGGAGAGAACACAGCACTTCTGGAGCGCACAGAGA 210
 Db 120 CCCAGTCTCTGGGTATAGAGTGTGAGGAGAGAACACAGCACTTCTGGAGCGCACAGAGA 179
 QY 211 CCGTGAAGATTCGAAGTTCAGAGAACTGCACCTTGGATGCCAAGATGCTTGAAGAAC 270
 Db 180 CCGTGAAGATTCGAAGTTCAGAGAACTGCACCTTGGATGCCAAGATGCTTGAAGAAC 239
 QY 271 AGCAGCCGAGCGGAGGCGCTCTCTGATGCTCCATGATCTTCAGCTCAGAAATCCT 330
 Db 240 AGCAGCCGAGCGGAGGCGCTCTCTGATGCTCCATGATCTTCAGCTCAGAAATCCT 299
 QY 331 GGATGAGAGCATCCCAAGGAGGAAATACCATCATGCTTGAATGCTTGAAGCCCATCG 390
 Db 300 GGATGAGAGCATCCCAAGGAGGAAATACCATCATGCTTGAATGCTTGAAGCCCATCG 359
 QY 391 GACTACTTGCACAAACACAGACACCGAGTGAACATCTGGGCTTTTCTCTGATATACCTT 419
 Db 360 GACTACTTGCACAAACACAGACACCGAGTGAACATCTGGGCTTTTCTCTGATATACCTT 419
 QY 451 TTCCGCGGCTTCTCTCTGAGGCGGCTGAGGCGGCGCACAAGAGGGGAGCTCTCAATGAAGA 510
 Db 420 TTCCGCGGCTTCTCTCTGAGGCGGCTGAGGCGGCGCACAAGAGGGGAGCTCTCAATGAAGA 479
 QY 511 CGTGTGTCTCTCTCTGAGCAGACAGTCTTCTGACGTGAACCTCAGAAAGACTAGAGACT 570
 Db 480 CGTGTGTCTCTCTCTGAGCAGACAGTCTTCTGACGTGAACCTCAGAAAGACTAGAGACT 539

| QY | 571 | GTGCAACAAGAGCTGATGTAAGTTGGCCAGACGCTGCTTCCCTGCCAAGGCTTGTG | 630 |
|--|---|--|-----------------|
| Db | 540 | GTGGCAAGAAGAGCTGATGTAAGTTGGCCAGACGCTGCTTCCCTGCCAAGGCTTGTG | 599 |
| QY | 631 | GATCTTCGCCGACACGAGCTCATCTGTCATCGTGTGCCGATGATCAGCAGCTGGC | 690 |
| Db | 600 | GATCTTCGCCGACACGAGCTCATCTGTCATCGTGTGCCGATGATCAGCAGCTGGC | 659 |
| QY | 691 | TGGCTTCAGTGGACACGCTTCATGG 716 | |
| Db | 660 | TGGCTTCAGTGGACCAATTTTCAGG 685 | |
| RESULT | 5 | | |
| LOCUS | U66687 | 1206 bp | EST 13-JUL-1999 |
| DEFINITION | HS066687 Human Homo sapiens cDNA similar to ATP-binding cassette transporter, mRNA sequence. | | |
| ACCESSION | U66687 | | |
| VERSION | U66687.1 | GI:1906572 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo. | | |
| AUTHORS | 1 (bases 1 to 1206) | | |
| TITLE | Alliknes, R., Gerrard, B., Hutchinson, A. and Dean, M. | | |
| JOURNAL | Characterization of the human ABC superfamily: Isolation and mapping of 21 new genes using the expressed sequence tags database | | |
| MEDLINE | Hum. Mol. Genet. 5 (10), 1649-1655 (1996) | | |
| COMMENT | 97049974 | | |
| FEATURES | Contact: Dean, Michael | | |
| source | National Cancer Institute, NCI-FCRDC | | |
| | Friedrick, MD 21702, USA | | |
| | Email: dean@crfvl.ncifcrf.gov | | |
| | derived from EST: EST277145 | | |
| | Location/Qualifiers | | |
| | 1..1206 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone_lib="Human" | | |
| BASE COUNT | 273 a 284 c 275 g 362 t | 12 others | |
| ORIGIN | | | |
| Query Match | 13.0%; Score 628.6; DB 189; Length 1206; | | |
| Best Local Similarity | 93.6%; Pred. No. 4.1e-147; | | |
| Matches 711; Conservative 0; Mismatches 37; Indels 12; Gaps 5; | | | |
| QY | 4054 | TCAGCTACTCTGGAACCTTGATCTGGAAGTATGGAGATGGGGATTACTTCTCACTGGG | 4113 |
| Db | 1 | TCAGCTACTCTGGAACCTTGATCTGGAAGTATGGAGATGGGGATTACTTCTCACTGGG | 60 |
| QY | 4114 | GGAAAGGAGAGCTCTGTGCATAGTAGAGAGCCCTGCTCCGCCACTGTAAGATCTGATTTT | 4173 |
| Db | 61 | GGAAAGGAGAGCTCTGTGCATAGTAGAGAGCCCTGCTCCGCCACTGTAAGATCTGATTTT | 120 |
| QY | 4174 | AGATAGAGCCACAGCTGCCATGAGACACAGACAGACAGACTTATTGATTTCAAGAGACCATCCG | 4233 |
| Db | 121 | AGATAGAGCCACAGCTGCCATGAGACACAGACAGACAGACTTATTGATTTCAAGAGACCATCCG | 180 |
| QY | 4234 | AGAAGCATTTTGACAGCTGTACCACTGTCACATTTGCCCATTCGCTGCACACAGGTTCTAGG | 4293 |
| Db | 181 | AGAAGCATTTTGACAGCTGTACCACTGTCACATTTGCCCATTCGCTGCACACAGGTTCTAGG | 240 |
| QY | 4294 | CTCCGATAGATTTATGTGCTGGCCACAGGAGCAGGTGATGGAATTTGACACCCCATTCGCT | 4353 |
| Db | 241 | CTCCGATAGATTTATGTGCTGGCCACAGGAGCAGGTGATGGAATTTGACACCCCATTCGCT | 300 |
| QY | 4354 | CCTTTGTTCACACGACAGTTCCCATTTCTATGCCATGTTTGTCTGCTGCAGAGAACAGGT | 4413 |
| Db | 301 | CCTTTGTTCACACGACAGTTCCCATTTCTATGCCATGTTTGTCTGCTGCAGAGAACAGGT | 360 |

| | | | |
|------------|---|---|------|
| OY | 4414 | CGCTGTCAAGGGCTGACCTCCCTCCTGTGGACGAGTCTCTTTCTTTAAGCATTCGCAT | 4473 |
| Db | 361 | CGCTGTCAAGGGCTGACCTCCCTCCTGTGGACGAGATTCTTTCTTTAGAGCATTCGCAT | 420 |
| OY | 4474 | TCCCGTGCTGGGGGGGCCCTTATCGCGCTCTCAACGAACCTTGGCTTCTGCAT | 4533 |
| Db | 421 | TCCCTGCTGGGGGGGGG - CCTTATCGCGCTCTCAACGAACCTTGGCTTCTGCAT | 479 |
| OY | 4534 | TTTTATCTTTGCGACAGCAGTTCGGGATTTGGCTTGTGTTCACCTTTTAGGAGATGCAT | 4593 |
| Db | 480 | TTTTATCTTTGCGACAGCAGTTCGGGATTTGGCTTGTGTTCACCTTTTAGGAGAGTCAT | 539 |
| OY | 4584 | ATTTTGATTATTTGATTATTTATTCATATTCACGTAAACAATAATTACTTTTGTCTTAAT | 4653 |
| Db | 540 | ATTTTGATTATTTGATTATTTATTCATATTCACGTAAACAATAATTGTTTGTCTTAAT | 599 |
| OY | 4654 | TGCACCTCAAAAAGGCTTCAGGG - AACCGTPTTPTTAATTTGATCAGAGG----CCATPAA | 4707 |
| Db | 600 | TGCACCTCAAAAAGGCTTCAGGGGAACCGTNTTPTTAATTTGATCAGAGGGCCATATTAAT | 659 |
| OY | 4708 | TGAACCTTTATACGTGA - GCTATATCTATATATATATTCGTACATAG---CCTATATT | 4762 |
| Db | 660 | GAAAGCTTTATACGCTGATGAGCTATATCTATATATATATTCGTACATAGGCCCATATTTA | 719 |
| OY | 4763 | TACACTGAAAAATGTAACTGTTTATTTATTTAAATAA 4802 | |
| Db | 720 | CAGGTGAAAAATGTAAAGCNGTTTANNTTATTTATTTAA 759 | |
| RESULT | 6 | | |
| LOCUS | BE386272 | 573 bp mRNA EST 21-JUL-2000 | |
| DEFINITION | 60127348BF1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3614828 5', | | |
| ACCESSION | BE386272 | mRNA sequence. | |
| VERSION | BE386272.1 | GI:9331637 | |
| KEYWORDS | EST. | | |
| SOURCE | human. | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| TITLE | 1 (bases 1 to 573) | | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/. | | |
| COMMENT | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC/DCTP CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov Plate: LNCM217 row: 0 column: 21 High quality sequence stop: 573. | | |
| FEATURES | Location/Qualifiers | | |
| SOURCE | 1..573 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /clone_image="IMAGE:3614828" | | |
| | /clone_id="NIH_MGC_20" | | |
| | /tissue_type="melanotic melanoma" | | |
| | /lab_host="DH10B (phage-resistant)" | | |
| | /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: | | |
| | EcoRI; cDNA made by oligo-dT priming. Directionally | | |
| | cloned into EcoRI/XhoI sites using the following 5' | | |
| | adapter: GGCAAGAG(G). Size-selected >500bp for average | | |
| | insert size 1.8kb. Library constructed by Ling Hong in | | |
| | the laboratory of Gerald M. Rubin (University of | | |
| | California, Berkeley) using ZAP-cDNA synthesis kit | | |
| | (Stratagene) and Superscript II RT (Life Technologies)." . | | |
| BASE COUNT | 119 a 150 c 156 g 148 t | | |

ORIGIN

Query Match 11.8%; Score 570.4; DB 167; Length 573;

Best Local Similarity 99.8%; Pred. No. 1.5e-132; Mismatches 571; Conservative 0; Indels 1; Gaps 0;

```

OY 2685 TGGCATTCCTGTTATATATAGCCCTTTTCATGCTGAATAGTACGACACCGCTTCAGCA 2744
DB 2 TGGCATTCCTGTTATATATAGCCCTTTTCATGCTGAATAGTACGACACCGCTTCAGCG 61
OY 2745 CCGTGTGTTGAGTTACTGATCAAGCAAGGAGGAGCAACCACTGTGATCGAGGA 2804
DB 62 CCGTGTGTTGAGTTACTGATCAAGCAAGGAGGAGCAACCACTGTGATCGAGGA 121
OY 2805 ACGACACCTCGGTGAGTGCACAGATGAGAGCAATCTCATATGCAATGATGCGAGCA 2864
DB 122 ACGACACCTCGGTGAGTGCACAGATGAGAGCAATCTCATATGCAATGATGCGAGCA 181
OY 2865 TGTAGCCCTCCATGAGGAGTGCATGCTGAAGCCATGAGAGGATGCTTTG 2924
DB 182 TGTAGCCCTCCATGAGGAGTGCATGCTGAAGCCATGAGAGGATGCTTTG 241
OY 2925 TCAAGGACACGCTGAGAGCTTCCTCCGAGTGCATGACGAGCTTTTCCGAAGATCCTTC 2984
DB 242 TCAAGGACACGCTGAGAGCTTCCTCCGAGTGCATGACGAGCTTTTCCGAAGATCCTTC 301
OY 2985 GAAGCCCTATGAGATTTTTCACAGACCCCAAGAGAGATTCACAGAGTTTCCA 3044
DB 302 GAAGCCCTATGAGATTTTTCACAGACCCCAAGAGAGATTCACAGAGTTTCCA 361
OY 3045 AAGACATGATTAATGATGACGTCGCGCTCCCTTCCAGGCCAGATGTTTCATCCAGAACG 3104
DB 362 AAGACATGATTAATGATGACGTCGCGCTCCCTTCCAGGCCAGATGTTTCATCCAGAACG 421
OY 3105 TTAATCTGCTGTTCTTCCTGTTGGGAAATGATGACGAGAGTCTTCCGAGTCTTGG 3164
DB 422 TTAATCTGCTGTTCTTCCTGTTGGGAAATGATGACGAGAGTCTTCCGAGTCTTGG 481
OY 3165 CAGTGGGCGCCCTGTCATCTCTTTTCAGTCCGACATGCTTCACAGGATCTGATTC 3224
DB 482 CAGTGGGCGCCCTGTCATCTCTTTTCAGTCCGACATGCTTCACAGGATCTGATTC 541
OY 3225 GGGAGCTGAAGGCTCTGACATATCAGCGAG 3256
DB 542 GGGAGCTGAAGGCTCTGACATATCAGCGAG 573

```

RESULT 7
 LOCUS BG403411 679 bp mRNA EST 12-MAR-2001
 DEFINITION 602419089P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4526110 5',
 mRNA sequence.
 ACCESSION BG403411
 VERSION BG403411.1 GI:13296859
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 679)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: AFCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLAMI0433 row: a column: 23
 High quality sequence stop: 667.

FEATURES

source

1. 679

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4526110"

/clone_lib="NIH_MGC_93"

/tissue_type="transitional cell papilloma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."

BASE COUNT 157 a 178 c 171 g 173 t
 ORIGIN

Query Match 11.7%; Score 569.2; DB 153; Length 679;

Best Local Similarity 96.8%; Pred. No. 3.2e-132; Mismatches 666; Conservative 0; Indels 13; Gaps 8;

```

OY 3883 GAGATCACTGATATTTGGCTTGGCCAGCTCCGAAGCAACTCTATCTCTCAAGA 3942
DB 1 GAGATCACTGATATTTGGCT--GCCAGCTCCGAAGCAACTCTATCTCTCAAGA 58
OY 3943 GCCGGTCTGTTCACTAGTGCACGTCAGATCAAAATTTGACCCCTTCACACGATACCTGA 4002
DB 59 GCCGGTCTGTTCACTAGTGCACGTCAGATCAAA-TTGACCCCTTCACACGATACCTGA 117
OY 4003 AAGACAGATTTGGGATGCCCTGGAGAGACACACATGAAAGATGATGCTCAGCTACC 4062
DB 118 AAGACAGATTTGGGATGCCCTGGAGAGACACACATGAAAGATGATGCTCAGCTACC 177
OY 4063 TCTGAACCTTGATGATGAGTATGAGAGATGAGAGATGAGTATCTCTCAGTGGGAAAGCGCA 4122
DB 178 TCTGAACCTTGATGATGAGTATGAGAGATGAGAGATGAGTATCTCTCAGTGGGAAAGCGCA 237
OY 4123 GCTCTTGTGATAGCTAGTAGAGCCCTGCTCCGCACTGTATGATTTGATGATGAGC 4182
DB 238 GCTCTTGTGATAGCTAGTAGAGCCCTGCTCCGCACTGTATGATTTGATGATGAGC 297
OY 4183 CACAGCTGCTATGAGACACAGACAGACACTTATGATTCAGAGAGACATCCGAGAAAGCAT 4242
DB 298 CACAGCTGCTATGAGACACAGACAGACACTTATGATTCAGAGAGACATCCGAGAAAGCAT 357
OY 4243 TGCAGACTGTACCATGCTGACCATGCCCCTGACACAGGTTCTAGGCTCCGATAG 4302
DB 358 TGCAGACTGTACCATGCTGACCATGCCCCTGACACAGGTTCTAGGCTCCGATAG 416
OY 4303 GATTTAGTGTCTGCGCCAGGAGACAGAGTGTGAGTTGACACCCCATGCGTCTCTGTC 4362
DB 417 GATTTAGTGTCTGCGCCAGGAGACAGAGTGTGAGTTGACACCCCATGCGTCTCTGTC 474
OY 4363 CAACGACATTTCCCATTTTATGCAATGTTTGTCTGCTGACAGAGACAGAGTGTGCA 4422
DB 475 CAACGACATTTCCCATTTTATGCAATGTTTGTCTGCTGACAGAGACAGAGTGTGCA 534
OY 4423 GGGCTGACTCTCCCTGCTGAGCAAGTCTTTTCTTTAGACATGTCATTCCTGCT 4482
DB 535 GGGCTGACTCTCCCTGCTGAGCAAGTCTTTTCTTTAGACATGTCATTCCTGCTGCT 594
OY 4483 GGGGGGGGGCCCTTCAATGCGTCTCTCAAGCAAACTTCTCTTCTGATTTATCTTT 4542
DB 595 TGGGGGGGGGCC-TCATTCGCT-CTCCATACGAAA-CTTGCTTCTCTGATTTATCTTT 651
OY 4543 CGCAGACAGCTTCCGATTTGCTTGTGT 4570
DB 652 CGCAGACAGCTTCCGATTTGCTTGTGT 679

```

RESULT 8

| LOCUS | BF816900 | 607 bp | mRNA | EST | 13-JAN-2001 |
|---------------------------|--|-------------|------|-----|-------------|
| DEFINITION | M3-C10184-081200-002-f10 C10184 Homo sapiens cDNA, mRNA sequence. | | | | |
| ACCESSION | BF816900 | | | | |
| VERSION | BF816900.1 | GI:12153115 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| AUTHORS | 1 (bases 1 to 607) Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. | | | | |
| TITLE | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags | | | | |
| JOURNAL | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) | | | | |
| MEDLINE | 2020263 | | | | |
| COMMENT | Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL. (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&t2=MR3-C10184-081200-002-f10&t3=2000-12-08&t4=1) Seq primer: puc 18 forward High quality sequence stop: 588. Location/Qualifiers 1..607 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="C10184" /dev_stage="Adult" /note="Organ: colon; ins: Vector: puc18; site:1: Smal; site:2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." | | | | |
| FEATURES | source | | | | |
| BASE COUNT | 131 a 166 c 161 g 148 t 1 others | | | | |
| ORIGIN | | | | | |
| Query Match | 11.6%; Score 564.6; DB 169; Length 607; | | | | |
| Best Local Similarity | 98.1%; Pred. No. 4.4e-13; | | | | |
| Matches 581; Conservative | 0; Mismatches 10; Indels 1; Gaps 1; | | | | |
| Y 2797 | TCGAGGGAAGAGACCTGCTGATGACAGCATGAAGACATCTCATATGACAGTACTA 2856 | | | | |
| Db 4 | TCGAGGGAAGAGACCTGCTGATGACAGCATGAAGACATCTCATATGACAGTACTA 63 | | | | |
| Y 2857 | TGCAGCATCTACGCCCTCTCCATGGCAGTCATGCTGTAAGACCATTCATATGAGAGT 2916 | | | | |
| Db 64 | TGCAGCATCTACGCCCTCTCCAGGAGCATGATGCTGATCTGAAAGCATTCGAGAGAT 123 | | | | |
| Y 2917 | TGCTCTTCTCAAGGCGACAGCTGCGAGCTTCCCTCCCGGCTGCATGACAGCTTTCCGAAG 2976 | | | | |
| Db 124 | TGCTCTTCTCAAGGCGACAGCTGCGAGCTTCCCTCCCGGCTGCATGACAGCTTTCCGAAG 183 | | | | |
| Y 2977 | GATCCTTCGAAGCCCTATGAGATTTTTCAGACAGACCCACAGGAGAGATTCACAG 3036 | | | | |
| Db 184 | GATCCTTCGAAGCCCTATGAGATTTTTCAGACAGACCCACAGGAGAGATTCACAG 243 | | | | |
| Y 3037 | GTATTCCAAAGACATGATGATGAGTTGACGTGGCGCTCCGCTTCAGAGCCGAGATGTTCA 3096 | | | | |

[illegible]

```
/strain="CB"  
/db_xref="taxon:9031"
```

| DEFINITION |
|--|
| ox08b10.x1 Soares_fetal_liver_spleen_1NF5_S1 Homo sapiens cDNA clone IMAGE:1655707 3' similar to SW:YL15_YEAST P14772 PROBABLE |

| | | | |
|----|------|--|------|
| Db | 181 | CTCTAAACTTGAATCTCAATGATGGAGAAATGGGGATTAATCTTCTCAATGGGGGAACGGC | 240 |
| OY | 4122 | AGCTCTTGTGCATAGCTATAGAGCCCTGCTCCGCCACTGTAAATTTCTGATTTTATGATGAAG | 4181 |
| Db | 241 | AGCTCTTGTGCATAGCTATAGAGCCCTTACTCTCCGTCACATGTAAAGATTCTGATTTCTAGATGAAG | 300 |
| OY | 4182 | CCACAGCTGCCATGGACACAGACAGACAGACTTATGATTTCAAGAGACCAATCCGAGAGAGAT | 4241 |
| Db | 301 | CCACAGCTGCCATGGACACCGAGAGACTTACTATCATCAAAAGACCAATCCGAGAGAGAT | 360 |
| OY | 4242 | TTCGACACTGTACCATCGTGTACCATTTGCCCAATCCGCTGTACACAGGTTTCTAGAGCTCCGATA | 4301 |
| Db | 361 | TTCGCTGATGTGACCACTGTGTGAAGATTTGGCCCAATCCGCTGTACACAGTCTTAGGCTCTGATA | 420 |
| OY | 4302 | GGATTAATGTGCTGTGCCACGAGGACAGGTGTGTGAATTTGACACCCCAATCGGTCTTCTGT | 4361 |
| Db | 421 | GGATTAATGTGCTGTGCCACGAGGAGGTGTGTGAATTTGACACCCCAATCGGTCTTCTTT | 480 |
| OY | 4362 | CCAAAGACAGTTCCCGATTTATGTGCATGTTTGGCTGTGC | 4401 |
| Db | 481 | CCAAAGACAGCTCCGGTTCTATGTGCATGTTTGGCTGTGC | 520 |

| RESULT | 13 | | | | |
|------------|------------|--------|-----------------|----------|----------------|
| BE749199 | BE749199 | 559 bp | mRNA | EST | 15-SEP-2000 |
| LOCUS | 199039 | MARC | 4BOV Bos taurus | CDNA 5', | mRNA sequence. |
| DEFINITION | BE749199 | | | | |
| ACCESSION | BE749199.1 | | | | |
| VERSION | | | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| CCW. | | | | | |

| ORGANISM | REFERENCE |
|---|---|
| Bos taurus | Smith, T. P. L., Casas, E., Stone, R. T., Heaton, M. P., Grosse, W. M., |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | |
| Bovidae; Bovine; Bos. | |
| 1 (bases 1 to 559) | |

| | |
|---------|--|
| TITLE | Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle |
| JOURNAL | Unpublished (2000) |
| COMMENT | Contact: Smith TPL |

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 107 row: O column: 23
Seq primer: ATTTAGGTGACACTATTAAG.

| FEATURES | source | Location/Qualifiers |
|------------|---|---------------------|
| | 1. .559 | |
| | /organism="Bos taurus" | |
| | /db_xref="taxon:9913" | |
| | /clone_lib="MARC 4BOV" | |
| | /tissue_type="pooled" | |
| | /lab_host="DH10B" | |
| | /note="Vector: pCMV SPORn6; Site_1: XbaI; Site_2: XhoI" | |
| | Library made from pooled tissue from day 20 and day 40 | |
| | embryos." | |
| BASE COUNT | 125 a | 118 g 171 t |
| ORIGIN | 145 c | |

| | | | | |
|-----------------------|--------|---------------------|---------|-------------|
| Query Match | 9.4%; | Score 454.4; | DB 139; | Length 559; |
| Best Local Similarity | 92.9%; | Pred. No. 2.5e-103; | | |

| | | | | | | | | | |
|---------|------|--------------|----|------------|-----|--------|----|------|----|
| Matches | 520; | Conservative | 0; | Mismatches | 36; | Indels | 4; | Gaps | 4; |
|---------|------|--------------|----|------------|-----|--------|----|------|----|

| | | | |
|----|------|--|------|
| QY | 4202 | GAGACAGACTATTGATTTCAAGAGACCATTCCGAGAAAGCATTTTGGCAGACTGTACCAATGCTG | 4261 |
| Db | 2 | GAGACAGACTTACTGA-TCGAGAGACCATTCCGAGAAAGCATTTGCTGTACTGCACACCAATGCTG | 60 |
| QY | 4262 | ACCAATGGCCCATCGCCGCGACAGGGTCTTAGGGCTCCGATAGAGATTATGGTGTGGCCAG | 4321 |
| Db | 61 | ACGATTTGCCCATGGCCCTACACACAGTCTTAGGCTCTGATAGGATTATAGTGTGGCCAG | 120 |
| QY | 4322 | GGACAGGTGGTGAGATTGACACCCCATCGGTCCTTCTGTCCACGACAGTTCGCCGATTTC | 4381 |
| Db | 121 | GGGACAGGGTGGTGAGATTGACACCCCATCGGTCCTTCTTCCACAGACAGCTCCCGGTTTC | 180 |
| QY | 4382 | TATGCCATGTTTGTCTGCTGCAG-AGAACAAGTTCGCTGCMAAGGGCTACTCTTCC-CTG | 4439 |
| Db | 181 | TATGCCATGTTTGTCTGCTGCAGAGAACGAAGTTCGCGCTCAAGGGCTTAACTCCCCCAATG | 240 |
| QY | 4440 | TTGACGAAAGTCTCTTCTTTTAGAGCATTCACATTCCTCGCCGCTGGGGGCGGGCCCTTCAT | 4489 |
| Db | 241 | CCGTTCACATCTCTTTCTTTTGGAGCATTCGCCATTCCTCTGCGGGCGGGCCCC-TCAC | 299 |
| QY | 4500 | CGCGTCTCTCAACGAAACCTTGCCCTTTCGATTTTATCTTTTCGACAGCAGTTCGGA | 4559 |
| Db | 300 | TGGCGTCTCTCGCCGAAACCTTGCCCTTTCGATTTTATCTTTTCGACAGCAGTTCAGGA | 359 |
| QY | 4560 | TTGGCTTGTGTGTTCACTTTAGGAGAGACTATTTTGGATTATGTATTATTCATA | 4619 |
| Db | 360 | TTGGCTTGTGTGTTCACTTTAGGAGAGACTCATTTTGGATTATGTATTATTCATA | 419 |
| QY | 4620 | TTCAATGTAACAAATTTAGTTTGTCTTATTTGACACTTAAAGGTTTCAGGGAACG | 4679 |
| Db | 420 | TTCAATGTAACAAATTTAGTTTGTCTTATTTGCACTTAAAGGTTTCAGGGAACG | 479 |
| QY | 4680 | TTATTTATPATTTGTATCAGAGGCTATPATGAACCTTTATACGTAGCTATPATATA | 4739 |
| Db | 480 | TTATTTATPATTTGTATCAGAGGCTATPATGAACCTTTATACGTAGCTATPATATA | 539 |
| QY | 4740 | TATTTCTGTATCATAGCCTAT | 4759 |
| Db | 540 | TATTTCTGTATCATAGCCTAT | 559 |

| RESULT 14 | | | |
|------------|------------|------------|---|
| AM655995 | AM655995 | 529 bp | mRNA |
| LOCUS | 108013 | MARC | IBOV Bos taurus cDNA 5', mRNA sequence. |
| DEFINITION | AM655995 | | |
| ACCESSION | AM655995.1 | GI:7421830 | |
| VERSION | EST. | | |
| KEYWORDS | | | |
| SOURCE | COW. | | |

| ORGANISM | TITLE | REFERENCE | JOURNAL | COMMENT |
|--|--|--|--------------------|--------------------|
| Bos taurus | Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle | Smith,T.P.L., Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid,W.W. and Keefe,J.W. | Unpublished (2000) | Contact: Smith TPL |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. | | 1 (bases 1 to 529) | | |
| Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahnenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid,W.W. and Keefe,J.W. | | | | |

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAACGCTATGACCAT

BACKWARD: GTTTCCAGTCACAGC
 Plate: 92 row: 1 column: 18
 Seq primer: ATTAGTGACACTATAG.
 Location/Qualifiers
 1. 529

FEATURES
 source
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 180Y"
 /issue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 112 a 147 c 116 g 154 t
 ORIGIN

Query Match 9.2%; Score 444.2; DB 119; Length 529;
 Best Local Similarity 92.4%; Pred. No. 9.1e-101;
 Matches 489; Conservative 0; Mismatches 38; Indels 2; Gaps 2;

QY 4136 GCTAGAGCCCTGCTCCGACCTGATGATTTAGATGAAGCCACAGCTCCATG 4195
 ||||||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 2 GCTAGAGCTTACTCCGCACTGTAAGTCTGATCTAGATGAAGCCACAGCTCCATG 61
 QY 4196 GACACAGACAGACTTATGATTCAGAGACCATCCGAGAAGCATTTGCACTGTACC 4255
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 62 GACACCGGAGACAGACTTACTGATCCAGAGACCATCCGAGAAGCATTTGCTAGCTGCAC 121
 QY 4256 ATGCGACCATTTGCCATCGCTGACACAGGTTGCTAGGCTCCGATAGGATTAATGTTGCTG 4315
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 122 ATGCTGAGATTTGCCATCGCTGACACAGTCTTCTGATAGATTAATGTTGCTG 181
 QY 4316 GCCAGGAGACAGTGTGATTTGACACCCCATCGTCTCTGTCCACAGACAGTTCC 4375
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 182 GCCAGGAGGAGGATGTGATGATTCGACACCCCATCGTCTCTTCCACAGACAGTCC 241
 QY 4376 CGATTTCTATGCTGATTTGCTGCTGACAGAAACAAGTCTGTAAAGGCTGATCTCTC 4435
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 242 CGGTTCTATGCAATGTTGCTGCTGACAGAAACAAGTCTGCTCAAGGCTTAACCTCCC 301
 QY 4436 C-CGTGACAGCAAGTCTTTCTTTTATAGCATTCGATCCCTGCGGGGCGGGCCCC 4494
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 302 CAGTCCGCTCCACATCTTTTCTTTTGAAGCATTCCTCCCTGCGGGGCGGGCCCC 361
 QY 4495 TTCTACGCTCTCTCTACCGAAACCTTCCTTCGATTTTATCTTTCGACACAGATT 4554
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 362 -TCACTGCTCTCTCTGCGAAACCTTCCTTCGATTTTATCTTTCGACACAGATT 420
 QY 4555 CCGGATTTGGCTTGTGTGTTTCACTTTAGGAGATCATTTTATGATTAATTTATTT 4614
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 421 CAGGATTTGGCTTGTGTGTTTCACTTTAGGAGATCATTTTATGATTAATTTATTT 480
 QY 4615 CCATTTATGATTAACAAATTTAGTTTCTTTTATGATTCGATTCCTAA 4663
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 481 CCATTTATGATTAACAAATTTAGTTTCTTTTATGATTCGATTCCTAA 529

RESULT 15
 AM761864 584 bp mRNA EST 04-MAY-2000
 LOCUS AM761864
 DEFINITION ug26f04.y1 Ren Stubbs mouse thymus Mus musculus cDNA clone
 IMAGE:2803615 5' similar to TR:014517 014517 SMRP.; mRNA
 sequence.
 ACCESSION AM761864
 VERSION AM761864.1 GI:7693777
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 584)
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
 ,B., Swaller,T., Gibbons,M., Pape,D., Harey,N., Schurk,R., Ritter
 ,E., Kohn,S., Sain,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterson,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MG1:1043139

FEATURES
 source

1. 584
 /organism="Mus musculus"
 /strain="C3H"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:2803615"
 /clone_lib="Ren Stubbs mouse thymus"
 /sex="mixed"
 /dev_stage="3 weeks"
 /lab_host="DH10B"
 /note="Organ: thymus; Vector: p7T73D-Pac; Site_1: NotI;
 Site_2: PacI; 1st strand cDNA was primed with an oligo(dt)
 primer; double-stranded cDNA was ligated using 5' linker
 ggcggcat and 3' linker aactggaacttaatt. Library is
 size-selected >2.5 kb and average insert size is 3.5 kb.
 Clones were arrayed from primary plating; non-amplified.
 Library constructed by X. Ren and L. Stubbs (Lawrence
 Livermore National Laboratory and DOE Joint Genome
 Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
 BASE COUNT 158 a 126 c 166 g 133 t 1 others
 ORIGIN

Query Match 9.0%; Score 434.6; DB 120; Length 584;
 Best Local Similarity 85.0%; Pred. No. 2.5e-98;
 Matches 498; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

QY 2271 TGGGCAACCATCTTCAATAGTGTATCCGGAACATCTCAAGTCAAGACGTTCTGT 2330
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 1 TGGGCAACCATCTTCAAGTGTATCCGGAACGCTCAAGTCAAGACGTTCTGT 60
 QY 2331 TTGTTACCCACAGTTACAGTACCTGTTGACTGATGATGATCTTCAATGAAGAGG 2390
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 61 TTGTTACACACAGTTACAGTATCTGTCGATGTGATGAGTGTATCTTCAATGAAGAGG 120
 QY 2391 GCTGTATTACGGAAGAGGACCCATGAGGAGATGATTAATTAATGATGATGCTA 2450
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 121 GCTGTATACAGAGAGAGTACCCATGAGGAGCTGATTAATTAATGAGGATTAACGCTA 180
 QY 2451 CCATTTTAAATTAACCTGTTGCTGGAGAGACACCGCGAGTGAATCAATCAATCAAAAAAGG 2510
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 181 CGATTTTAAATTAACCTGTTGCTGGAGAGACACCCCGAGTGAATCAATCAATCAAAAAAGG 240
 QY 2511 AAACAGTGTGTTACAGAGAAGTACAGAAGCAAGGTTCTTAAACAGAGCAATTAAGA 2570
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 241 AAGCTACTGTGTTTAC-AAAATATCAAGACAAAGGCCCTTAAGCCAGGTACAGTAAGA 297
 QY 2571 AGGAAAAGACAGTAAGCCAGGAGGAGGAGGAGCTTGTGCAGCTGGAAGAAAAGGCGAGG 2630
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 298 AGGAGAGAGGCGGTGAAGTCGAGGAGGAGGAGGAGCTTGTGCAGGTGAGAGAAAAGGCGAGG 357
 QY 2631 GTTCAAGTCCCTGCTGATATATGTTGTCTACATCCAGGCTCTGCGGGCCCTTGGAAT 2690
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 358 GTTCTGTGCTGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
 QY 2691 TCCGTGTTATTAATGAGCCCTTTTCATGTAATGATGAGCAGACCGCCCTTACAGACCTGGT 2750

```

Db 418 TCCGTGATCATGATGTCCTCTTCATGCTGAATGTGGGAGACACTGCTTCAGCACCCTGCT 477
QY 2751 GGTGAGTTACTGATCAAGCAAGGAGGAGACACCACTGTGACTGAGGGAACGAGA 2810
Db 478 GCTTATCTACTGACCAAGCAAGGAGGAGACAGCAGCAGTGTATCAATGGAACAGAA 537
QY 2811 CCTCGTGTGACAGCANTGAAGCAATCCCTCATATGCGACTACTA 2856
Db 538 NCTTCGTGAGTGACAGCATGAATGACAGCCCTTCATGCGAGTACTA 583

```

Search completed: November 9, 2001, 14:03:15
 Job time: 8563 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: November 9, 2001, 15:51:42 ; Search time 73.13 Seconds
(without alignments)
1191.258 Million cell updates/sec

Title: US-09-528-031-2

Perfect score: 7308
Sequence: 1 MKDDIGREYIIIPSPGYSV.....DSSRFYAMFAAENKVAWG 1437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601: *
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseq/AA2001.DAT: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 7308 | 100.0 | 1437 | 21 | AA10225 |
| 2 | 7304 | 99.9 | 1437 | 19 | AA10225 |
| 3 | 7298 | 99.9 | 1437 | 22 | AA10225 |
| 4 | 7293 | 99.8 | 1437 | 20 | AA10225 |
| 5 | 2094.5 | 28.7 | 1541 | 18 | AAW3361 |
| 6 | 2082 | 28.5 | 1531 | 19 | AAW57486 |
| 7 | 2082 | 28.5 | 1531 | 20 | AAW9894 |
| 8 | 2082 | 28.5 | 1531 | 20 | AAW74471 |
| 9 | 2082 | 28.5 | 1531 | 21 | AAW78873 |
| 10 | 2082 | 28.5 | 1531 | 21 | AAW5799 |
| 11 | 2082 | 28.5 | 1545 | 18 | AAW3362 |

| | | | | | | |
|----|--------|------|------|----|---------|--------------------|
| 12 | 2075.5 | 28.4 | 1489 | 17 | AA10225 | Multi-drug resista |
| 13 | 2075 | 28.4 | 1531 | 15 | AA10225 | Multi-drug resista |
| 14 | 2075 | 28.4 | 1531 | 17 | AA10225 | Multi-drug resista |
| 15 | 2075 | 28.4 | 1531 | 19 | AA10225 | Multi-drug resista |
| 16 | 2075 | 28.4 | 1531 | 20 | AA10225 | Multi-drug resista |
| 17 | 2075 | 28.4 | 1531 | 20 | AA10225 | Multi-drug resista |
| 18 | 2075 | 28.4 | 1531 | 21 | AA10225 | Multi-drug resista |
| 19 | 2075 | 28.4 | 1531 | 21 | AA10225 | Multi-drug resista |
| 20 | 2075 | 28.4 | 1531 | 21 | AA10225 | Multi-drug resista |
| 21 | 2070 | 28.3 | 1545 | 19 | AA10225 | Multi-drug resista |
| 22 | 2055.5 | 28.1 | 1528 | 19 | AA10225 | Multi-drug resista |
| 23 | 2055.5 | 28.1 | 1528 | 20 | AA10225 | Multi-drug resista |
| 24 | 2055.5 | 28.1 | 1528 | 20 | AA10225 | Multi-drug resista |
| 25 | 2055.5 | 28.1 | 1528 | 21 | AA10225 | Multi-drug resista |
| 26 | 2055.5 | 28.1 | 1528 | 21 | AA10225 | Multi-drug resista |
| 27 | 2049.5 | 28.0 | 1525 | 20 | AA10225 | Multi-drug resista |
| 28 | 2006.5 | 27.5 | 1482 | 17 | AA10225 | Multi-drug resista |
| 29 | 1993.5 | 27.3 | 1622 | 19 | AA10225 | Multi-drug resista |
| 30 | 1977 | 27.1 | 1621 | 19 | AA10225 | Multi-drug resista |
| 31 | 1941.5 | 26.6 | 1528 | 18 | AA10225 | Multi-drug resista |
| 32 | 1941 | 26.6 | 1527 | 20 | AA10225 | Multi-drug resista |
| 33 | 1861.5 | 25.5 | 1472 | 17 | AA10225 | Multi-drug resista |
| 34 | 1828 | 25.0 | 1355 | 21 | AA10225 | Multi-drug resista |
| 35 | 1807.5 | 24.7 | 1352 | 21 | AA10225 | Multi-drug resista |
| 36 | 1807.5 | 24.7 | 1511 | 21 | AA10225 | Multi-drug resista |
| 37 | 1778.5 | 24.3 | 1545 | 19 | AA10225 | Multi-drug resista |
| 38 | 1769.5 | 24.2 | 1215 | 21 | AA10225 | Multi-drug resista |
| 39 | 1755 | 24.0 | 1147 | 21 | AA10225 | Multi-drug resista |
| 40 | 1749 | 23.9 | 1212 | 21 | AA10225 | Multi-drug resista |
| 41 | 1745.5 | 23.9 | 1580 | 19 | AA10225 | Multi-drug resista |
| 42 | 1743 | 23.9 | 1772 | 21 | AA10225 | Multi-drug resista |
| 43 | 1743 | 23.9 | 1776 | 21 | AA10225 | Multi-drug resista |
| 44 | 1743 | 23.9 | 1787 | 21 | AA10225 | Multi-drug resista |
| 45 | 1730 | 23.7 | 1477 | 18 | AA10225 | Multi-drug resista |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AA10225 | standard; Protein: 1437 AA. |
| ID | AA10225 |
| AC | AA10225 |
| DT | 16-NOV-2000 (first entry) |
| DE | Human MRP-beta protein. |
| XX | |
| KW | MRP-beta; multidrug-resistance associated protein; human; cytostatic; |
| KW | chemotherapy; cancer cell; drug discovery; cytotoxic. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | US6077936-A. |
| XX | |
| PD | 20-JUN-2000. |
| XX | |
| PF | 16-APR-1998; 98US-0061400. |
| XX | |
| PR | 16-APR-1997; 97US-0843459. |
| XX | |
| PA | (MILL-) MILLENNIUM PHARM INC. |
| XX | |
| PI | Shyjan A; |
| XX | |
| DR | WPI: 2000-430613/37. |
| XX | |
| DR | N-PSDB: AAA40481. |
| XX | |
| PT | Model multidrug-resistance associated polypeptide useful for improving |
| PT | the effectiveness of a chemotherapeutic regimen to eradicate |
| PT | multidrug-resistant transformed cells especially cancer cells - |

PS Claim 1: Fig 1A-D; 43pp; English.
XX
CC This invention describes a novel model multidrug-resistance associated
CC polypeptide, MRP-beta (I) which has cytostatic activity. Inhibitors of
CC (I) are useful for improving the effectiveness of a chemotherapeutic
CC regimen to eradicate multidrug-resistant transformed cells, especially
CC cancer cells, from the body of a mammal, preferably human. (I) is also
CC useful for drug discovery, especially to the design of novel
CC chemotherapeutic drugs that are cytotoxic to cells expressing (I). This
CC sequence represents the human MRP-beta protein which is described in the
CC method of the invention.
XX
SQ Sequence 1437 AA:

Query Match 100.0%; Score 7308; DB 21; Length 1437;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDIDIGKEYITPSGYSVREKSTSTGTHREDSEKFRRTPLKCODALETAARAGLS 60
DB 1 mkdldlgkeyilpspyrsrterstsgthredskfrtrtpcqdaletaaraegls 60
QY 61 LQASMSQRLIDBEHPKGYHNGLSALPRTYCKQHGVNNAGLFSCWTFESWLSLAR 120
DB 61 ldsmsqrlildeehpkyghnglsalprtyckqhgvdnaglfscwtfeswlsar 120
QY 121 VAKKGGELMEDWLSKSHSSDVNCRRLERLMOEELNENYGPDAASLRVYVICRRLI 180
DB 121 vahkggelsmedwlskshssdvncrrlerlwgeelneynpdaaslrvyvicrrli 180
QY 181 LSTVICMTTOLAGFSGPAFWKHLLEYQATESNLQYSLLVGLLTLTETIVRSWLA 240
DB 181 lsvicmtltqlagfsgpafwnkhlleytqatesnlqyslllvgl lltetivrswl 240
QY 241 ALMYRQVRLRGALITMAFKKILKLIKESIGELINICSDNGRFEAAVSSLAG 300
DB 241 almyrtvrlrgalitmakfkilklkikesigelinicsndgrfearvsslaag 300
QY 301 PVAALIGMIVNIIIGPTGFGSAVFILFYPAWMAFASRLRAYFRKCAVATDERVOKME 360
DB 301 pvaaligmivniiigptgfgsavgfilyfypawmafasrlrayfrkcaaatdevokme 360
QY 361 VLYYIKRKMYAWKAFSOSVOKIREERRILEKAGYFQSIIVGAPIVVIAVVFESV 420
DB 361 vlyyikrkmyawkafsqsvokireerrilekagfyqsiivgvapivviasvffsv 420
QY 421 HMTLGRPLTAAQAFVTVVNSMTFALKVTPFSYKLSSEASVADRKSFLMEVHM 480
DB 421 hmtlgrpltaaqafvttvnsmtfalkvtpfsykslseasvadrksflmeevhmik 480
QY 481 NKRPASPIKTEKKNATLAMPSSSHSSINSPKLPKMKKDRASGKKEKROLDRT 540
DB 481 nkpspiktekknatlawpssshssinspklpkmmkdrasgkkekroldrt 540
QY 541 VLAEGQKHLILDSDERPSPREEGKATHLGLKQRTLSIDLEIOGKLVGICG 600
DB 541 vlaegqkhlildsderpspeegkathlglkqrtlshidleiegklyvgicgsv 600
QY 601 KTSLSIALIGOMTLESTASISGFAYVAQAAWTLNLTLDNLLFGKEVDEER 660
DB 601 ktslslaligomtlestaslsigfayvaqawtlndnllfgkevdeern 660
QY 661 CCLRPDLAIPSSDLTEIGEGANLSCGORISLARALYSDRSITLIDPLSLDA 720
DB 661 ccrlpdlai pssdlteigeranlsgqgrislarsl ysdrsiylldplslda 720
QY 721 NHTFNSAIRNHLKSTVLFTVHQLVDCDEVLFEMKEGCIITERGTHEELM 780
DB 721 nhtfnsairnhlkskvtlftvqhqlvdcdevlfmkegciitergtheelmn 780
QY 781 FNNLLGEPVPELINSKKEKSGSKSQDQPKTGSIKKKAVPPEGQVLQLEK 840
DB 781 fnnllgepvpelinskkesgsksqdqpkgtgskkavppegqvlqllek 840

DB 781 fnnllgepvpelinskkesgsksqdqpkgtgskkavppegqvlqllek 840
QY 841 VPMSVGVYVIOAGCPALFVITMALFMLINVGSTAFSTWMLSTYKIOGSGNTTV 900
DB 841 vpmsvgyvyioagcpalfvitalfmlinvgstafstwmlstykiogsgnttv 900
QY 901 VSDSMKQNPMMQYVYASITAYALSMAVMLLKAIKGVYFVKGLIRASSRL 960
DB 901 vsdsmknpmmqyvyasitayalsmavmllkaikgvyfvgklirassr 960
QY 961 KEFDTPTGRTINRFSKDMQEVDRLPFOAEMFQNTIVLFFCGMAGVPMFLV 1020
DB 961 kfdtpgtgrtinrfskdmqevdrpfoaemfqn timerlffcgmagvpmflv 1020
QY 1021 LVILFVLIHIVSRVILRELKRLDINTQSPFLSHITSSIOGLATIHAANKGO 1080
DB 1021 lvilfsvlihvsvr ilrelkrl dntqspflshitssio glatihaynkgo 1080
QY 1081 LDDNQAPPELFTCAMRWLAVRDLISALTITTTGLMTVLMHGOIPPAVAG 1140
DB 1081 lddnqapfflftcamrwlvavrdlislaltttglmtvlmhgoipavag 1140
QY 1141 GLPQFYRLASEPFAKTSVRINHYIKTSLKAPRIKKNAPSPDPOGSEVFE 1200
DB 1141 glpftyrlasepfaktsvrinhyikt slkaprikknapsdpogsevfena 1200
QY 1201 RYRENLPLVLKRVFTYIKPEKIGIVRTGSGKSSLGMAFLRVELSGGCIK 1260
DB 1201 ryrenlplvlk rvyftykpekigivrtgsgksslgmalrvelsggcik 1260
QY 1261 IGLADRSKLSITPOEVLFSGVTRSNLDPNOYTEDQIWDALERTHMKECIA 1320
DB 1261 igladrsklsitpoevlfsgvtrsnldpnoytedqiwdalertmkeci 1320
QY 1321 SEVENGDNSVGEKROLCTARALLRHCKILLIDETAAADTEFDLLIOETIR 1380
DB 1321 sevengdnsvge krolctarallrhckillidetaaadtefdllioetir 1380
QY 1381 MLTFAHRLHTVLSGDRIMVLAQGVFEFDPVSLNDSRFFYAMFAAENK 1437
DB 1381 mltfahrlhtvls gdrimvlaqgvfefdpvslndsrrff yamfaaenk 1437

RESULT 2
ID AAM80597 standard; Protein: 1437 AA.
XX
AC AAM80597;
XX
DT 02-FEB-1999 (first entry)
XX
DE Human multidrug resistance-associated protein.
XX
KW Multidrug resistance-associated protein; MRP-beta; human; cancer;
KW tumour; chemotherapy; therapy.
OS Homo sapiens.
OS
PN WO9846736-A1.
PN
PD 22-OCT-1998.
PD
PF 16-APR-1998; 98WO-0507673.
PF
PR 16-APR-1997; 97US-0843459.
PR
PA (MILL-) MILLENNIUM PHARM INC.
PA
XX Shyjan A;
XX
PI
PI
DR WPI: 1998-568724/48.
DR N-PSDB; AAV65682.
XX

PT New isolated multidrug resistance-associated polypeptide - used to
PT develop products for modulating multidrug resistance, particularly
PT for reducing resistance of tumours to chemotherapeutic drugs

PS Claim 26; Page 54-59; 93pp; English.

XX This is the amino acid sequence of novel human multidrug resistance
CC associated protein (MRP-beta), as deduced from an isolated cDNA
CC clone (see AAV65682). Over-expression of MRP-beta is thought to be
CC associated with the emergence and/or persistence of a multidrug-
CC resistance phenotype in transformed mammalian cells, including
CC carcinoma and adenocarcinoma cells. MRP-beta is probably a
CC transmembrane-spanning, energy-dependent transporter or pump that
CC functions by ejecting intracellular substrates. The invention
CC provides compositions and methods for improving the effectiveness
CC of chemotherapeutic regimens to eradicate multidrug-resistant
CC transformed cells from the body of a mammal, especially from the
CC body of a human. The disclosed compositions include MRP-beta
CC nucleic acids, including probes and antisense oligonucleotides
CC (see also AAV65684-88). MRP-beta polypeptides and antibodies,
CC MRP-beta expressing host cells, and non-human mammals that are
CC transgenic or nullizygous for MRP-beta. The disclosed methods
CC include methods for attenuating aberrant MRP-beta gene expression,
CC protein production and/or protein function, and for improving the
CC effectiveness of chemotherapy for a mammal afflicted with a
CC multidrug-resistant tumour, wherein the tumour is of mammary,
CC respiratory tract, urogenital tract, endocrine system or immune
CC system origin. In addition, methods are disclosed for identifying
CC and using a modulator, such as an inhibitor, that is cytotoxic to
CC cells expressing MRP-beta.

XX Sequence 1437 AA:

Query Match 99.9%; Score 7304; DB 19; Length 1437;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDIDIGEYIIPSPGYSVRERTISTSGTHRDREDSKFRFRPLECODALETAAEGIS 60
DB 1 mkddidigeyiipspgyvsrvtertsqthrdredskfrfrplcqdalaetaaegis 60
QY 61 IDASMSQRLRIDEBHPGKGYHHGSLAKPIRTCKHOHPVDNAGFSCMPTSMISLAR 120
DB 61 idasmsqrlriidebhpqkykhhgslakpirtckhqpvdnaglfscmtswislar 120
QY 121 VAHKKGELSMEDVWSLSKRESSDVNCRLERIWQELNEVGPDAASLRVYVIFCRTL 180
DB 121 vahkkgelmedvwslskressdvncrrlerlwgeelnevgpdaaslrvvwifcrtl 180
QY 181 LSIYGLMTTQLAGFSGPAFWKHLLEYQATFSNIOYSLILVGLILLETIVYWSGLATW 240
DB 181 lsiyglmttqlagfsgpafwmkhlleyqatesnlqysllvlgllleivyswslatw 240
QY 241 ALNRYTGVRLGALITMFAFKLILKLNKEKSLGELINCSNDGQRMFPAAVGSLAG 300
DB 241 alnrytgvrlgalitmfafklilknikekslgelincsndgqrmfpaavgsllag 300
QY 301 PVVALIGMIVYVNIIGPFGUGSAVILFYPAAMPASRLTAYFRKKCYAATDERVOKNE 360
DB 301 pvvaligmivyniigpfgugsavilfypammpasrltayfrkkcyaatdervokne 360
QY 361 VLTYYIKFTKMTAWKAFESQSQKIREERRILIEKAGYFQSTIVGAPIVYVYVTSY 420
DB 361 vlttyikftkmtawkafesqsqkireerriliekagyfstivgapiyyvvyvtsy 420
QY 421 HMTGFDLTAAQAFVYVYVNSMTPALKVTFPSVKSLSSEASVAVDFRSLFMEEVHMK 480
DB 421 hmtgfdltaaqafvyvvyvnsmtfalkvtfpsvkslseasvavdfrrslfmeevhmk 480
QY 481 NKPASPHRIEKKNATLWMDSSHSSIONSPIRTPMKKDKRASRCKKEKRVOLQRTHQ 540
DB 481 nkpasphrikknatlwmdsshssionspirtpmkkdkrasrckkekrvolqrthq 540

QY 541 VLAEOGHLIDSDRSPSEEEKHINHLRLORTLSIDLEIOEKLIVGSGVSG 600
DB 541 vlaeoghliidsdrspseeeqghnhlrlortlsidleioegklyvgsgvsg 600
QY 601 KTSLSIALIGQWTLLEGSLAISGTFAYVAQOAMINATLRDNLPGKEDEERYSVLNS 660
DB 601 ktslsialigqwtllegslaisgtfayvaqawlratlrndllpgkeydeerysvlns 660
QY 661 CCLRPDLALPSSDLTEIGERCANISGGORORISARALYSRSTIYLDDPLSALDAHV 720
DB 661 cclrpdlalpssdlteigercanissggororisaralysrstiylddplsaldahv 720
QY 721 NHIFSAIRKHLKSTVLFVTHQLQYVDCDEVITMKKECTERGTHERMLNDGYATI 780
DB 721 nhifsairkhlkstvlfvthqlqyvdcddevitmkectergtthermlndgyati 780
QY 781 FNNLLIGETPVEINSKETSQSKSODKGPRTGSIRKEKAVKEEGQLVLEEKQGS 840
DB 781 fnnlligetpveinsketssqsksqdkgprtgssirkeekavkeegqlvleekqgs 840
QY 841 VPMVYGYVYIQAGCPPLAVLIMALFMLNVGSTAFSTWMLSTWIKQSGNTVTNGNETS 900
DB 841 vpmvygyvyiqagcpplavlimalmfmlnvgstafstwmlstwikqsgntvtngnets 900
QY 901 VSDSKDNPMMQYASISYALSMANVMLIKAIRGVYFVGTLAASSRLDELFRRLRSPM 960
DB 901 vsdskdnppmmqyasisyalsmanvmlikairgvvyfvgtlaassrldelfrlrrspm 960
QY 961 KPEDTPGRIINRFSKMDDEVDRLPFOAEMFIONVILVFCVCMINGVFPFPLVAVCP 1020
DB 961 kpedtpgriinrfskmddevdrlpfoaemfionvilvfcvcmingvfpfplvavcp 1020
QY 1021 LVILFVSALIVSRVILRELKRLDNTQSPFLSHITSIOGLATIAVYKQGEFLHRYOEL 1080
DB 1021 lvilfvsalivsrvilrelkrl dntqspflshitsioqlatiahvaykqgeflhryel 1080
QY 1081 LDDNAPPEFLFCAMRWLAVRDLISIALITTTGIMIVLMHQIQPAXAGLISVAVOLT 1140
DB 1081 lddnappeflfcamrwlvavrldlisialitttgimivlmhqiqpaxaglaisvavolt 1140
QY 1141 GLFOFTVRLASEEFAEFTSVERRINHYKITLSLEAPARIKNKAPSPDMPQEGEVTEENAE 1200
DB 1141 glfgftvrlaseeefstverrinhykitlsleapariknkapsdpmpqegevteenaem 1200
QY 1201 RYRENULPLYLKVSTFIKPEKIGIVRTSGKSSLGMAFLPVLDELGGCIIIDGVRISD 1260
DB 1201 ryrenulplylkvstfikpekigivrtsgksslgmaflpvlvelsggcikidgvrisd 1260
QY 1261 TGLADRSKLSITIQEPVLFSGTAVSNIDPFNQYTEDQIWMALERTHMEKICIAOLPIKLE 1320
DB 1261 tgladrsklsitiqepvlfsgtavnidpfnoytedqiwmalertmhmeckiaolplikle 1320
QY 1321 SEVMENGDNFVSGERQLCICARALRHCKILILDATAMDETETLLIOETIRFAFDCT 1380
DB 1321 sevmengdnfvsgerqlcicaralrhckilildatamdetetllioetirfafdct 1380
QY 1381 MLIARHLTVLGSORINWLAOGVVEEDFVSVLSDNSRFYAMPAAENKVAVK 1437
DB 1381 mliarhlvtvlgso rinwlaogvveedfvsvl sdnssrfyampaenkvaavk 1437
RESULT 3
AAB47021
ID AAB47021 standard; Protein: 1453 AA.
XX AAB47021;
XX 29-MAR-2001 (first entry)
XX Multidrug-resistance associated polypeptide-beta.
XX Multidrug resistance-associated polypeptide; MRP-beta; transporter;
KW

KM P-glycoprotein; chemotherapy; cancer.
XX OS Homo sapiens.
XX US6162616-A.
PN XX
PD 19-DEC-2000.
XX
XX 16-APR-1997; 970S-0843459.
XX PF
XX 16-APR-1997; 970S-0843459.
PR 16-APR-1997; 970S-0843459.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA ShyJan A;
XX PI
XX WPI: 2001-111728/12.
DR N-PSDB; AAC85287.
PS
PT Novel nucleic acid encoding multidrug resistance-associated protein
MP MRP-beta, useful for detecting and treating drug-resistant cancers -
PS
PS Disclosure; Fig 1; 41pp; English.
XX This sequence represents a multidrug resistance-associated
CC polypeptide (MRP-beta). MRP-beta is a transporter that can remove
CC chemotherapeutic agents from cells and/or sequester such agents,
CC and it probably accounts for multidrug resistant phenotypes that do
CC not respond to treatments with reversal agents directed against
CC P-glycoprotein or the known MRP. By inactivating MRP-beta, at nucleic
CC acid or protein levels, cells should be rendered sensitive to
CC chemotherapy agents. The MRP-beta coding sequence, or fragments
CC of it, can be used to detect expression of the MRP-beta gene and
CC thus multidrug resistant cells. Fragments of MRP-beta DNA are useful
CC as primers and as antisense (therapeutic) agents for inhibiting
CC expression of the MRP-beta gene. They can also be used for detecting
CC mutations and to treat diseases associated with abnormal MRP-beta
CC gene expression, specifically cancer, particularly for potentiating
CC chemotherapy of fluid or inoperable tumors, or in cases of regrowth
CC after an initial course of chemotherapy.
XX
XX Sequence 1453 AA:
SQ
Query Match 99.9%; Score 7298; DB 22; Length 1453;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1436; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MKQIDIGKEVITIPSPGTSVRERKSTSGTHRDSDSKFRRTREPCDALTARAGLS 60
DB 17 mkdidiqkylipspgyrsvrertstsgthrdedskfrrtrplecdalataaagls 76
QY 61 LDASMHQSLRLDDEHNGKGYHNGLSALPKRTCKHOHPDNAGLFCSCMFSMLSLAR 120
DB 77 ldsamhbsqlrllddehnpkykhynglsalpkrrtsknhpndnaglsfcmfslslar 136
QY 121 VAHKGGLSMEDVNSLSKHSSSDVNCRLERLMOBELNEVGPDAASLRVWICRTRLI 180
DB 137 vahkgglsmedvnslskhsdvnrcrlrlmoeelnevgsdaaslrwvwcrtll 196
QY 181 LSTVCLMITQLAGSGAFAMWKHLLETQATSNLOYSLLVGLLLETYSRSLATW 240
DB 197 lsvclmitqlagsgafamwkhllletqatesnlqyslllvglletyrsvaltw 256
QY 241 ALNRTGVRRLGATLTAAFKILTKNIKESLDELINICSDGORFEAAVAGSLAAG 300
DB 257 alnrtgvrirlgatltaafkilknikeslgellnicsdgrmfdaavagslaag 316
QY 301 PVVAILGMIVNIIIGPTGLGSAVFTLFFPAMMFASRLTAYFRRKCVAAATDEVOKNE 360
DB 317 pvvailgmivniiigptglgsavftlffpammfasrltayfrkkcaatdevokne 376
QY 361 VVIYIKIKIKAMVAKAFSOSVOKITREBRILKEAGVFQSTTGVAPLIVVIVASVTF 420

DB 377 vlyikikikmawkatsvqkireerillkaagyfgstlvqvapilvvlaavtfsv 436
QY 421 HMTLGFDLTAQAFTVTVTVNSMTFALKVTPFSYKSLSEASVAVDRKSLFLEMEVMIK 480
DB 437 hmtlgfdltaqaftvtvtnsmtfalkvtpfsykslseasvavdrkslfleemvmmik 496
QY 481 NKPASPHIKTEMKVATLAODSSSHSIONSPKLVKMKDKKRASKGKKKRYOLORTFHOA 540
DB 497 nkpasphiktemkvatlawdssshsionspkltvkmkdkkraskgkkyqlqrteha 556
QY 541 VLAOKGHLLDSDRSPPEEKGKHLHGLRLQRTLSHIDLEIOGKLVGTGSVSG 600
DB 557 vlaeqlhllsderspeeegkhlhglrlqrthshidleiqeqlvgjcsvsg 616
QY 601 KTSLSAILGOMTLIEGSIASIGTFAYVAQAWILNATLRDNLIFGKEYDEERYNSYLS 660
DB 617 ktslailgomtlliegslasigtfayvaqawilnatlrnllifgkeydeerynsyls 676
QY 661 CCLRPDLALPSSDLTEIGERGANLGGQORISLARALYSDRSIYLLDPLSLADAHV 720
DB 677 cclrpdlalpsdltteigeranlsgqqrqrlsalaralysdrsiyllddplsaldahv 736
QY 721 NHITNSAIRKHLKSKTYLPTVHOLOYVDODEVTFMKEGCTTEGTHBELMNLNGVATI 780
DB 737 nhitnsairkhlksktylptvlhqlqylvdcdevlfmkegcttegtlhelmmngvati 796
QY 781 FNNILLGETPPEVENSKRKETSQSKSODKPKTKSIRKEKAVRPEGOLVOLBEKSGS 840
DB 797 fnnillgetpvelnskrketsqsksgkdkpkrtslkkkavppegqlvqleekggs 856
QY 841 VPSVYGYVIOAAGPLAFLYALMALFNLNGSTAFSTWMLSYTKQSGNTVTRGNETS 900
DB 857 vpsvygyvioaagplaflyalmalfnlmgstafstwmlsytkqsgntvtrgnet 916
QY 901 VSDMKNPQHGYASYATASMAVMLIKATRGVFPKGTARASSRLHDEFRILLSPM 960
DB 917 vsdmknphgyasyatasmavmlikatrkvfpkgtrassrlhdelfrillsp 976
QY 961 KFFDTTPGRILNRFPSKDMDEVDRALPFOAEMFIONVILVFFCGMILAGVPMFLVAVGP 1020
DB 977 kffdtptgrilnrfpskdmdevdrvlpfoaemfionvllvfcgmlagvpmflvavgp 1036
QY 1021 LVILFSLVLIHVRVLRLEKRLDNITQSPFLSHITSSIQGLATTHAYNKGQELHRRQEL 1080
DB 1037 lvilfsvlihsrvlrlkrlnditqspflshitssiqglathaynkgeflhrygel 1096
QY 1081 LDDNOAPELFTCAMRLAVRDLISALTITGMLVIMHGOIPPAVAGIATSYAVOLT 1140
DB 1097 lddnapelfltcamravlrdlislaltitglmvlmhgoipparavagiatyavolt 1156
QY 1141 GLFOFTVARLASETARFSTVSRINHYIKTSLLEAPARIKKNAPSDMPOEGEVTFENAE 1200
DB 1157 glfotvarlasetarfstvsrinhyiktlslleaparikknapsdmpogevtfenaem 1216
QY 1201 RYRENPLVLKKSFTTKPKKIGIVGRTSGKSLGMALFRLVELSGGCIKIDGVRI 1260
DB 1217 ryrenplvlkksfttkpkkigivrtsgkslgmalfrlvelsggckidgvrisd 1276
QY 1261 IGLADRSKLSITPOEPLFSGVRSNLDPNQOTEOIDALERTHMKECIALDPLKLE 1320
DB 1277 igladrsklsitpoeplfsgvrsnldpnqotoidalrthmkecialdplkle 1336
QY 1321 SEVENMGDNFSGEROLCTARALRLRCKLILDEATRAMDTELDLLOETIRAFADCT 1380
DB 1337 sevenmgdnfsgerqlctaralrlrcklildeatramdtedlldetirafadct 1396
QY 1381 MLTIAHRLHTVAGSDRIWVLAQGVVEFDPVSVLLSNDSSRFYMPAAENKVAVK 1437
DB 1397 mltiahrlhtvagsdriwvlaqgvvefdpvsyllsndssrfymfaaenkvavk 1453

RESULT 4

AA43542
ID AAY43542 standard; Protein: 1437 AA.
AC AAY43542;
XX 26-JAN-2000 (first entry)
DE A human MPR-related ABC transporter designated MOAT-C.
XX Human; MPR-related ABC transporter; MOAT protein; MOAT-C;
KW MOAT mediated transport; anticancer drug sensitivity;
KM transporter mediated cellular efflux; anticancer.
XX Homo sapiens.
OS
XX MO9949735-A1.
FN
PD 07-OCT-1999.
XX 26-MAR-1999; 99WO-US06644.
PE
XX 27-MAR-1998; 98US-0079759.
PR
PR 03-AUG-1998; 98US-0095153.
XX (FOX-C-) FOX CHASE CANCER CENT.
PA
XX Krush G, Lee K, Belinsky M, Bain L;
PI
XX MPI: 1999-610812/52.
DR
DR N-PSDB: AA230079.
XX
PT New transporter gene useful for screening for anti-cancer drugs -
XX
XX Claim 16; Page 135-138; 153pp; English.
XX
XX The present sequence represents a human MPR-related ABC transporter
CC (MOAT) protein, designated MOAT-C. The protein comprises a multi-domain
CC structure including a tandem repeat of nucleotide binding folds
CC appended C-terminal to a hydrophobic domain, having Walker A and B ATP
CC binding sites and several potential membrane spanning domains. The MOAT
CC nucleic acids are useful for screening a test compound for inhibition of
CC MOAT mediated transport, indicated by restoration of anticancer drug
CC sensitivity, which in turn causes a reduction of transporter mediated
CC cellular efflux of anticancer agents. MOAT DNA or RNA may be used as
CC probes to detect the presence or expression of genes encoding MOAT
CC proteins. Anti-MOAT antibodies are useful for detecting and quantitating
CC MOAT proteins.
CC
XX
XX Sequence 1437 AA:
SQ
Query Match 99.8%; Score 7293; DB 20; Length 1437;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1434; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 241 alnytrgyrlrgaalltmafkklklnklsjgelnlncsdngqrmfleaavgsllag 300
QY 301 PVVAILGMITVYVILIGPTGFGSANYFIIFYPAMMFASRLTAFFRKCYAADDERQKANE 360
Db 301 pvaallgmlyvillpgcflgsavfllfypammfarsltafyfrkcvaaaterqkme 360
QY 361 VLTYSFKFKMYAWKAFQSOYKIREERIRLEKAGYFQSTIVGAPVIVVIVASVTFVS 420
Db 361 vlytklklmyawvkaifsgvqkltreeerllekagyfqlgtvgyvaplvvasvltsv 420
QY 421 HMTLGFEDTLAQAFTVWVFNFSMPFALVTPFSYKSLSEASVADRFKSLFLMEVHM 480
Db 421 hmtlgfdltaagafctvrcvfnsmfalkvtpfsvkslseasvavdrfkfslfmeevhmk 480
QY 481 NKPSAPHIKIEKKNAVTLAMDSHSSIQNSPKLTPMKKDKRASRCKEKERYOLOPTEHOA 540
Db 481 nkpsaphikieknatlawdsahssisqnspltkmkdkdkastrgkkekvrqlgrtlehpa 540
QY 541 VLABQKHLLDSDERPPEEEGKHHLGHRLQRTLSIDLEIOEKLKVGICGSVGS 600
Db 541 vlaeqkghllldsderpspeeeqkhhghlgrlrlhsldeleqgklvqicgsvgs 600
QY 601 KTSLSAILGQWTLLEGSIASGTFAYVAQAQAMINATLRDNIILRGKEXDERYSVLS 660
Db 601 ktslsailgqwtlllegsaistgftayvaqgawlnatlrhnlilgkkeydeerysvls 660
QY 661 CCLRPDLAIPSSDLTEIGERGANSQGORISLARALYSDRSITYLDDPVSALDAHVG 720
Db 661 cclrpdlalpsdlteligeranlsqgrqrlsalaralsdrrsyllddpvsaladahvg 720
QY 721 NHIFNSAIRKHLKSTVLFYFHQLOLYVDCDEVIFPMKECCTERGTHERMLNDVYATI 780
Db 721 nhifnsairkhlkstktylftvqhlylvdcdevilmkegccltergltheelmldngdyati 780
QY 781 FNNLLIGETPPVEINSKRETSQOKKSDKCPKTSIKKEKAVKEBEOQLVLEEKGGCS 840
Db 781 fnnlligetppveinskretsqokksdqkspktsyvkkekvpeegqlvgleekgqgs 840
QY 841 VPMVSYGYVYIQAGGPLAFLVIMALEFNLVNGSTABSTWMLSWIKOGSGNTVTNGETS 900
Db 841 vpmvsygyvyiqaggpflafvlmalflnvgstabstwmllswikogsgntvttrgnet 900
QY 901 VDSKKNDRPHQYASTYALSAWMLILKATRGVYFVNGTLAASSRLDELFRRLRSRM 960
Db 901 vdsdkndrphmqyasialsmawmlilkatrgvfvngtlaassrlndelfrlrllrsm 960
QY 961 KFEFTTPGRILNRFSSKMDDEVRLPROAEMFIONVILVFCVGMINGVFPMPFVAVGP 1020
Db 961 kfedttppgrilnrfsskmddevrllproaemfionvillvfcvgmingvfpmpfivaavp 1020
QY 1021 IVILFSLHIVSRVLIRELKRLDNTOSPFSLHITSIOGLATIHAYNKGOFELHRYOEL 1080
Db 1021 ivilfslhivsrvlirelkrldntospfslhittsioqlatihaynkogeflhyrge 1080
QY 1081 LDDNQAPEFLPTCAMRWLAVALDLISIALITTTGMLTYLMBQIPRAYAGLAISTAVQLT 1140
Db 1081 lddnqaepflftcamrwlavrlidlisialitttgmltylmbqiprayaglaistavqlt 1140
QY 1141 GLFOFTVRLASFEARFVSVERINHYIKTSLSEAPARKNKPSDMPDQEGEVTEPNEEM 1200
Db 1141 glfoftvrlasfearfvsverinhyiktlsleaparknkpsdmpdqpgegvteneem 1200
QY 1201 RYRENLPVLKKVSEFTIRPEKEIGIVRTSGKSSLGNALFRLVELSGGCIRIDGVRISD 1260
Db 1201 ryrenlpvlkkvseftirpekigivrtsgksslgmalfrlvelsggcikidgvrissd 1260
QY 1261 IGLADRSKLSITPOEPVLFSGTVSNLDPFNQYTEDQIMDLERTHMEKCAQAPKLE 1320
Db 1261 igladrsklsitpoevulfsgtvsnldpfnqytedqimdlertmhmeckaqpikle 1320
QY 1321 SEVMENGNDFSVGEROLCIARALRHCKIILDEATAMDEFTDLIOETFREAFADCT 1380
Db 1321 sevmengndfsvgerqlliciaralrhckilldeataamdetdliogelireafadct 1380

AAW57486 standard; Protein; 1531 AA.

| | |
|---|--|
| AC | AAM57486; |
| XX | 14-AUG-1998 (first entry) |
| DT | |
| XX | |
| DE | Human MRP variant ltpga (lei/pgpa). |
| XX | |
| KW | Multidrug resistance-associated protein; MRP; tumour; human; variant; |
| KM | multidrug resistance; MDR; leishmania P-glycoprotein; ltpga; lei/pgpa. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FH | Key Location/Qualifiers |
| FT | Misc-difference 685 /label= L685S |
| FT | /note= "wild-type Leu is replaced by Ser" |
| FT | Misc-difference 1282 /label= R1282A |
| FT | /note= "wild-type Arg is replaced by Ala" |
| PN | US5766880-A. |
| XX | |
| PD | 16-JUN-1998. |
| XX | |
| PE | 05-JUN-1995; 95US-0463092. |
| XX | |
| PR | 05-JUN-1995; 95US-0463092. |
| PR | 27-OCT-1992; 92US-0966923. |
| PR | 08-MAR-1993; 93US-0029340. |
| PR | 26-OCT-1993; 93US-0141893. |
| PR | 20-MAR-1995; 95US-0407207. |
| XX | |
| PA | (TOOH) UNIV QUEBENS KINGSTON. |
| PI | Cole SP, Deeley RG; |
| XX | |
| DR | MP: 1998-361687/31. |
| NFSDR: | AAV31498. |
| XX | |
| DNA encoding protein associated with multi-drug resistance - useful for as probe for identifying multi-drug resistant tumour cells | |
| PS | Claim 1; Columns 67-78; 82pp; English. |
| CC | This represents a variant of the human multidrug resistance-associated protein (MRP). This natural variant is a leishmania P-glycoprotein related molecule ltpga (lei/pgpa). The human and murine MRP nucleic acid molecules can be used as probes for identifying multidrug resistant tumour cells by hybridisation to mRNA from tumour cells. The antisense nucleic acid can be used to reverse multidrug resistance (MDR). A recombinant expression vector containing the MRP nucleic acid molecules operatively linked to at least one regulatory sequence can be used to transform a host cell to produce a recombinant MDR-associated protein. |
| SO | Sequence 1531 AA; |
| Query Match | 28.5%; Score 2082; DB 19; Length 1531; |
| Best Local Similarity | 34.5%; Pred. No. 1,1e-165; |
| Matches 488; Conservative 262; Mismatches 479; Indels 184; Gaps 21; | |
| OY | 100 PVDNAGLSCMTFSWLSSILARYAHKKGELSMEDVSLSHESDVNCRRRLERLMQEEINE 159 : : : : : : : : : : : : : : Db 209 pssasfslrlftwltgll-vgyrfgplsgsdwlshkedsqvvlpvvlvkhwkecak 267 OY 160 V-----GPPDA-----SLKRVWFCTRLL 181 : : : : : : : : : : : : : : Db 268 trkgpvkvvysskdapagkesskydaneaveallvkspdkewmpslfkylktfigpyflm 327 OY 182 SLVCLMIQLGLGFSPAPMAVKHLLEYQTATESNLGYSLLVLGILLTEIRVMSLATMA 241 328 sffkaahldlmfmfsqpq-ilkllikfvndckapdwggyfytyvlllvtaciqtlvhqyfh 386 DY 242 INFRGVLRGAAILMARKKIIKLNIKEKS--LGEIINICSNDOGRPFEAAYVSLLAG 299 |


```

QY 160 V-----GPDAA-----SLRKYVATFCTRLL 181
Db 268 trkqpvkvyskddpapkesskvaneveallvkspqkewmpslfvylykftgpyfm 327
QY 182 SYVCLMTIQLGEGSPAPMVKHLEETQATESNLQYSLLVGLLTFEIVRMSLALPMA 241
Db 328 sfffaialdmfmspgp-ikllikfyndtkapdqgyfyfyllvtacqlvlhygfh 386
QY 242 LNYRTGVRRLGAILMARKIKLIKNIKES--LGLJINICSDQORMEEAAGVSLLAG 299
Db 387 lcfvsgmrklcavlgayrkalvlnsarkstvgelivlnmsvdqgrlmdlatylnmlws 446
QY 300 GPVVAIILCMITNVILGFTGFLGSVFLTFPAMMFASRLAYFRKCYAATDEVQKKN 359
Db 447 aplqylalalyllwlnlpsvlagavmwlmvvnvnammkcktygvalmkkskdnriklmn 506
QY 360 EVLTYIKFKIKYAWKAKSQSVOKIREEBRILEKAGYFOSITVGAVPIVYVIAVTFPS 419
Db 507 eilngikvikiyawaiaakdvaiatrgaelkvikksaylsagvcltwtcplvalccta 566
QY 420 VHMFLGPD--LTAQAFTVTVVFNMSFPALKVTPSVKSLSEASVAVDRFKSLFMEVH 477
Db 567 vyvliidennlidaqtafvsalfoilfrplnlpmvissivagvsalkrltriflshoeel 626
QY 478 --MKKNKASP---HIKEMKNATLAMDSSHSSSIONSPKLPFKMKKDKRASKREKVRQ 532
Db 627 pdslerfpvkdggsfnslvratnltw----- 653
QY 533 LQRTGQAVLAEQKCHLLDSDERPSEEBGKHILGHLRLQRTLSIDLEIOBKLVG 592
Db 654 -----arsdpp-----tlngitfsipealva 675
QY 593 ICGSVSGKTSLSIALIQMTLLGSGISALSGTFAYAAQAMILNATLRDNLIFGEKYEDE 652
Db 676 vvgvgcgkssllsallaeamdvgehvaikgvayvpgawqndslrenliffgqleep 735
QY 653 RYNSVLNCCCLRPDLALIPSSDLTEIGERGANTLSCGORISIALALSKDSRITLIDPL 712
Db 736 yrrsvlgacallpdllelpsqdrteigevnlsgqgkqvslayavsnadlylfdpdl 795
QY 713 SALDAHVNHTIFNSAI--RKHLKSKTVLFTVHQLOYLVDCCDEVIFMKEGCTERGTHEEL 770
Db 796 savdahvgkhlfenvlpgkmlkxktrllvthsmyslpqvadliymsgklsiemsgyel 855
QY 771 MNLNGDYATIF-----NNLL-----GETPPVEIN 795
Db 856 lardgafaelrtyasteqdaengvtvsgpgkaekqmengmvltdsaqklqqrqls 915
QY 796 SKKETSGSOKSODKGRPTGSIKKKAKVPREGOLVOLEKQSGSVPMVYGVYIOAGG 855
Db 916 ssssvsgdistrhn--staellqkaeakkeetwklmeadkaqtgvlsvywdynkaiaql 972
QY 856 PLAFVIALMFLMANGSTAFSTWMLSYWKQSGNTVTTCGNETSVSDSKDNPHMYXA 915
Db 973 fislflsif-lfmchvsalaasnyvlstldpdivngt---gehtkvrslsvygalaisggl 1028
QY 916 SIYALSMAMVILKAKRGVVFVKGTLRASSRLHDELFRRIILSPKFFDPTPTGRLNLF 975
Db 1029 avfgyssmavsl-----ggllasrcihvdlhslrpsmffertpsgnlvmrf 1076
QY 976 SKDMDEVNVRLEPQAEMLTQNVILVFCVGMTAGVFPWFVLAVGVLILFSLHLYSRVL 1035
Db 1077 skeldvtasmlpevlkmtmgslfnvlgacivillatpilaailipplgily---ffvqrfly 1133
QY 1036 I---RELKRLDITQSPFLSHITSSIOGLATTHAYNKGOEFLHARQOELDDNQAFLEFT 1092
Db 1134 vassrgklrlevsvrpyshnetllysvvtrateegerfihgsklxdenkqkypsi 1193
QY 1093 CAMRLAVRLDLISIALITTTGMLTVMHGOIPPAVYAGIAISYAVOLTGRLFQFTVRLASE 1152
Db 1194 vanrjavrlecvgncivlfaallfavlsrshsiasglvglsvsgyslqvtylwnlvmrsee 1253

```

```

QY 1153 TEARETVERINHYIKTSLSEAPRIKKNKAPSPDMPOEGEYTFEENAEKMYRENPLVLKK 1212
Db 1254 metulvaaverlkeyset-ekesapwqigetappsswpqygrvefrnycilrtyedldfvlrh 1312
QY 1213 VSFITKPKKEKIGIVGRTSGSKSLGMAFLVELSGGCIRKIDGVRISPDIGLADLSKLSI 1272
Db 1313 lnvtlungekgvlgvrtgagkscltclgfrinesaegeildglnlakglnldrfkltl 1372
QY 1273 IPOEPVLESGTVRSNLDPPNOYTEDQIWDALERTHMKECIAOLPLKLESEVMEKDNFSV 1332
Db 1373 lpgdvlftsgslrmldpfsygsdeevwtstelahlkdfvsalpkldhecegegenlsv 1432
QY 1333 GEROLCTARALLRCKIILIDEATAMDTERTDILQETIRAFADCMULTAHHLHPLV 1392
Db 1433 gqrqlvclatallrlkrlklivideataavldetdldlqslrtlrqfcdetcvlthnrlntlm 1492
QY 1393 GSDRIWVLAQGVFEFDPSPVLSLNDSSRFYAM 1425
Db 1493 dylrvlvdkgelgeigpapsdl-qgrqlfysm 1524

RESULT 8
AAW74471
ID AAW74471 standard; Protein: 1531 AA.
XX
AC AAW74471;
XX
DT 18-MAY-1999 (first entry)
XX
DE Human multidrug resistance-associated protein variant.
XX
KM Multidrug resistance-associated protein; MDR; human; diagnosis;
KM MDR tumour cell identification; cancer therapy.
OS Homo sapiens.
XX
PN US5882875-A.
XX
PD 16-MAR-1999.
XX
PF 05-JUN-1995; 95US-0462109.
XX
PR 05-JUN-1995; 95US-0462109.
PR 27-OCT-1992; 92US-0966923.
PR 08-MAR-1993; 93US-0029340.
PR 26-OCT-1993; 93US-0141893.
PR 20-MAR-1995; 95US-0407207.
XX
PA (TOOH ) UNIV QUEENS KINGSTON.
XX
PI Cole SPC; Deeley RG;
XX
DR WPI; 1999-214061/18.
XX
DR N-PSDB; AAX21977.
XX
PT Identifying a multidrug resistant tumour cell by contacting the cell
PT with an antibody/antigen-binding fragment - which binds to an
PT expressed protein encoded by multidrug resistance-associated protein
PT (MRP) nucleic acid
XX
PS Claim 3; Column 69-80; 80pp; English.
XX
CC This sequence is the human multidrug resistance-associated (MDR)
CC protein. The invention relates to a method for identifying a multidrug
CC resistant (MDR) tumour cell. Compositions and methods utilising the MDR
CC proteins can be used to treat patients with tumours displaying multidrug
CC resistance, particularly those displaying resistance to antineoplastic
CC epipodophylotoxins, vinca alkaloids, and hydrophobic drugs. The methods
CC for inhibiting/killing a MDR tumour cell can be useful for treating
CC breast cancer, leukemias, fibrosarcomas, cervical cancer, gliomas,
CC thymas, neuroblastomas and lung cancer. The MDR DNA sequences when
CC labeled are useful as molecular probes for diagnosing multidrug
CC resistance of a tumour (using cells from a tumour biopsy) and for

```


XX 01-APR-1998 (first entry)

XX Human canalicular multispecific organic anion transporter protein.

XX Canalicular multispecific organic anion transporter protein;

XX cMOAT protein; ATP-binding cassette transporter family; ABC transporter;

KW hepatobiliary excretion; multidrug resistance-associated protein;

KW cMOAT protein activity; multidrug resistance-related protein; MDR-1;

KW Dubin-Johnson disease; Rotor disease.

XX Homo sapiens.

OS

XX Key

XX Modified-site

XX Location/Qualifiers

XX 7..9

XX /note= "Asn is predicted to be N-glycosylated"

XX 26..49

XX /note= "putative transmembrane region"

XX 74..87

XX /note= "putative transmembrane region"

XX 106..118

XX /note= "putative transmembrane region"

XX 130..148

XX /note= "putative transmembrane region"

XX 166..184

XX /note= "putative transmembrane region"

XX 319..341

XX /note= "putative transmembrane region"

XX 361..379

XX /note= "putative transmembrane region"

XX 440..456

XX /note= "putative transmembrane region"

XX 463..480

XX /note= "putative transmembrane region"

XX 546..566

XX /note= "putative transmembrane region"

XX 583..605

XX /note= "putative transmembrane region"

XX 671..679

XX /note= "putative transmembrane region"

XX 761..771

XX /note= "Walker A, B or signature sequence"

XX 781..786

XX /note= "Walker A, B or signature sequence"

XX 976..996

XX /note= "Walker A, B or signature sequence"

XX 911..913

XX /note= "putative transmembrane region"

XX 1025..1050

XX /note= "Asn is predicted to be N-glycosylated"

XX 1101..1120

XX /note= "putative transmembrane domain"

XX 1022..1039

XX /note= "putative transmembrane region"

XX 1212..1232

XX /note= "putative transmembrane region"

XX 1238..1251

XX /note= "putative transmembrane region"

XX 1334..1342

XX /note= "putative transmembrane region"

XX 1437..1447

XX /note= "Walker A, B or signature sequence"

XX 1457..1462

XX /note= "Walker A, B or signature sequence"

XX /note= "Walker A, B or signature sequence"

XX W09731111-A2.

XX PD 28-AUG-1997.

XX 21-FEB-1997; 97WO-NL00079.

XX 22-FEB-1996; 96EP-0200460.

PA (MED-) ACAD MEDISCH CENT AMSTERDAM.

PA (HEFN-) HET NEDERLANDS KANKER INST.

PA (INTR-) INTROGENE BV.

XX Borst P, Bosma PJ, Evers R, Oude Elferink RPJ;

PI Paulusma CC;

XX WPI: 1997-435163/40.

DR N-PSDB: AAT94023.

XX DNA encoding human and rat canalicular multispecific organic anion

PT transporter proteins - useful for diagnosis and treatment of

PT Dubin-Johnson disease and Rotor disease

XX Disclosure: Fig 10; 106pp; English.

XX The present sequence represents a novel canalicular multispecific

CC organic anion transporter (cMOAT) protein. The protein is a new member of

CC the ATP-binding cassette (ABC) transporter family. The ATP dependent

CC cMOAT transporter system mediates hepatobiliary excretion in the liver.

CC cMOAT may be a liver-specific homologue of multidrug

CC resistance-associated protein. The nucleic acids are used to provide

CC cells with cMOAT protein activity. cMOAT protein activity in cells can be

CC enhanced by increasing the level of glutathione, glucuronide and/or

CC sulphate. Antisense constructs, especially derived from another multidrug

CC resistance (MDR)-related protein, e.g. MDR-1, to the nucleic acids and

CC vectors can be used to decrease the level of cMOAT in a cell. The nucleic

CC acids and proteins can be used especially in diagnosis of Dubin-Johnson

CC disease, Rotor disease or another disease involving cMOAT. The cMOAT gene

CC may also be used as a selectable marker gene.

XX Sequence 1545 AA;

SQ

Query Match 28.5%; Score 2082; DB 18; Length 1545;

Best Local Similarity 34.5%; Pred. No. 1.2e-165;

Matches 502; Conservative 246; Mismatches 488; Indels 218; Gaps 27;

QY 99 HPVDNAGLFSCMTFSWLSLARVAHKGELSMEDYWSLSKHESSDVNCRRLRMOEIN 158

DB 193 npsiasfssitlyswysillkykr-pltledvewdeemkktlvskfethmkrelq 251

QY 159 EV-----GPDASLRV-----VW-- 172

DB 252 karalqrrqekssqngsagarlpjlnknsgsqdaivledvkkkkksgckkdvpxswlm 311

QY 173 --IFCRTLIL--SIYCLMITQLAGSGAPFVKKHLETOATESNLOYSLLVGLILT 228

DB 312 kalftftymvllksfilrlklniditfvsbpq-llkllisfasdrdylywgyicalllfta 370

QY 229 EIVRSWSIALFWALNYRTGVRGLGAILTMAFKKILKNI--KEKSLGELINICSDGOR 286

DB 371 aligfcicqcyfqlcfklgkvkrtaimasyvkaltlnlarkeytvgectvlmsvdaqk 430

QY 287 MFEAAVGSILAGPVPVAILGMIVYVILGPTGFGSAVFILFYRAMFPASLTAYFRK 346

DB 431 lmdvnlfmhmlwssvlgjvlisflwrelgsvlgvymvvlvlnalstkskrlqvk 490

QY 347 CVAATDERVOKNEVLTYIKFKIMYAWVAKAFOSVOKIREERRLIEKAGYPOSTITGVA 406

DB 491 mnknkdkrikimellsgkiklyfawepstfdqynlkkaklnllaftsqdgcvvilvf 550

QY 407 PIVVIVASVTFESVHMTLGF--LTAAQAFVTVTVENSMTFALKYTPSVKSLSPASVAV 464

DB 551 qltpvlsvvtfsyvvlvdsnnldaqkafstltlfnlrfplmnpmmismlgasvat 610

QY 465 DRFKSLFLEEVYHMKTKNRPASPH-----IKTEKNATLAMDSSHSSIONSPLTLTKMKKD 519

DB 611 erlekylgddl----dtsalrhdcnfcfkamqfseastf----- 646

QY 520 KRASGKKKVKVLOLQTEHOAVLABOKGHLLDSDRSPSPREEGKHILHGLRLQRTLH 579

DB 647 -----ehds-----eatvr 655

```

OY 580 SIDLEIOEGKIVGCSVSGKSTLSIALIGOMTLLEGSLAISGTFAYVAQAAMILNATL 639
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 656 dvalldmagqlvaiviprvsgksllsamlgemenyghltltgtaayvqgswlqgtli 715
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 640 RDNILFEKEYDEERYNSVLSNCCRPDLALPPSDLTEIGERCANLSGGORISLARAL 699
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 716 kdmliftetnekrqgvleacallpdlmlpgsdlaelgekqinlsqgkqrlslarat 775
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 700 YSRRSYIIIDDPISALDAHGNHFNFAIRKH--LKSQTLFVTHOQYVDCDEVFMK 757
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 776 yqndlyllldqpsavdanhkhlnkvipgnllkqctllvchmhrlpydeivvlg 835
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 758 EGCITERGTHEELMNLNGDYATTFNNLLGCTPPE----- 793
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 836 ngtlvekgysallakkgefaknlkflrhtgpeeeetvhdgseeddyyllssveip 895
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 794 -----INSKETS-----GSQKKSQDKPKTGSi---KKEKAVKPEEGOLVQ 832
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 896 edaasltmrrensfrltlrssrsngrlhkslnsktrvnslkedeeilvkqg--klilk 953
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 833 LEEKGSGVPMWYGYVIOAAGGPIAFLVIMALEMNVGSTAR--STWMLSYWIKQSGN 890
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 954 kelfegkvkfslyleylqglfsifflila-fvnm--svafignslwlsaw----- 1003
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 891 TTYTRGNETSVSDSMKDNPMHOYYASIALMAVMILKAIRGV-VFVK-----GTLRA 943
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1004 -----tsdsklfn--stdypasqrdmrvgyvlglaqglfvfiahfwsafgtvha 1052
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 944 SSRULDELFRILIRSPKFFDDTPRTGRILNRFSDMDEVVRLEPQEMFIQNVILVFFC 1003
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1053 snllhqlnlllrprrtltctptgrivrtfagdstvtdlpqslrsvtclfgllst 1112
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1004 VGMIAGVPMFLVAVGVPLVLEFVLIHVSVLLIRELRLNITQSPFLSHITSIOGLAT 1063
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1113 lvmclmatpvtlivlplgllysvsgmfystsrqlrtldsvtclpshysfsevsqilpv 1172
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1064 IHAVNKGQELHRYQELLDNQAFPLEFTCAMRLAVRLDLISIALITTTGLMIVLMHQG 1123
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1173 lrtebhqgrflkhneerltdnqkcvfswltsnrvlairlvlgnlvffsalnmvlyrdt 1232
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1124 IPRVYGLAISYAVOLTGLQFYVRLASFEFARTSVBRINHYIKTISLEAPAIKKAP 1183
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1233 lsgdlvgyfvlsnlnltqlmwlvrmtseletnlvaverlleyk-venaeap-wvtdktp 1290
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1184 SPMPQGEYTFENAEKRYRENPLVLEKVSFTIKPEKIGIVRGTSKSGSLGMALFRL 1243
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1291 ptdkpskqkqifmnyvrrypeldlvtlrgltcdlsgmeklgyvgrtgaqksltnclfrl 1350
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1244 VELSGGCIKIDGVARISDGLADLRSLKSLIIPQEFVLFSGTVRSNLDPFNOYTEDQIDAL 1303
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1351 leaagqglldgydlaslgldlrkeltltpqdbllfsgslrlmldpfnmgsdeekwal 1410
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1304 ERHMRKCIQOLPLKLESEWENCDNSGEROLCTARALLRCKLIIIDEAANAADTE 1353
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1411 elallskfsvaslqglshvteaggnlslqgrllclgrillsktlivldeaaavdle 1470
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1364 TDLLOETIREAFADCTMLTIAHRLHVLGSDRLIMVLAQGVFEFDPVSLVLSNDS--RF 1422
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1471 tdnltgtitqnefahcvltiahrlhtlmsdkvmlvndngklileysgeepellqpgpfyf 1550
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 1423 YAMFAAENKVAVK 1436
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1531 makeaglenvstck 1544
    :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 12

AAR96952 ID AAR96952 standard; Protein; 1489 AA.

XX AC AAR96952;

XX "08-AUG-1996 (first entry)

```

XX DE Multi-drug resistance protein alternative splice form 1.
XX KM MRP: multi-drug resistance protein; small cell lung cancer; protect;
XX KM cytotoxicity; therapy; P-glycoprotein; alternative splice variant.
XX OS Homo sapiens.
XX FH Key location/Qualifiers
XX FT MISC-difference 154..155
XX PR /note="amino acids 155-196 of AAR93153 removed"
XX PN US5489519-A.
XX PD 06-FEB-1996.
XX PF 27-OCT-1992; 92US-0966923.
XX PR 26-OCT-1993; 93US-0141893.
XX PR 27-OCT-1992; 92US-0966923.
XX PR 08-MAR-1993; 93US-0029340.
XX PA (TOOH ) UNIV QUEENS KINGSTON.
XX PT Cole SPC, Deeley RG;
XX DR MPI; 1996-115615/12.
XX DR N-PSDB; AAT14910.
XX PT Multi-drug resistance protein, MRP and DNA encoding it - for
XX PT diagnosis and treatment of e.g. small lung cell cancers
XX PS Disclosure; Column -: 50pp; English.
XX CC The present sequence is that of an alternative splice form of a multi-
XX CC drug resistance protein (MRP). The specification states that nucleotides
XX CC 657-783 of AAT17173 (encoding full-length MRP) are removed (amino acids
XX CC 155-196 inclusive of AAR93153), however this does not maintain the
XX CC reading frame, hence nucleotides 658-783 were removed to create AAT14910
XX CC which encodes the present sequence. The deletion eliminates a potential
XX CC secretory signal cleavage site located between amino acids 189-190 of
XX CC AAR93153). The MRP gene was isolated from H69AR, a small cell lung
XX CC cancer cell line, which is about 50-fold resistant to doxorubicin as
XX CC compared to the parental H69 cell line. Immunogenic portions (see also
XX CC AAR96957-58) of the protein were also identified and of use in the
XX CC invention. The MRP protein is overexpressed in multi-drug resistant cells
XX CC independently of overexpression of P-glycoprotein (a member of a
XX CC superfamily of membrane proteins that serve to transport a variety of
XX CC mols., ranging from ions to proteins, across cell membranes, and
XX CC associated with a variety of tumours). The nucleic acid encoding MRP (see
XX CC AAT17173), and its splice variants, etc. can be used to protect a drug
XX CC sensitive cell from cytotoxicity due to exposure to a drug, or to develop
XX CC improved forms of therapy, particularly for small cell lung cancers which
XX CC exhibit a high initial response to chemotherapy which is almost always
XX CC followed by a multi-drug resistance form of the disease.
XX SQ Sequence 1489 AA:
OY 71 ILDEHPKGRKYNHGL-----SALKPRTYCKKHQ-----PVDNAGIFSCTFSWL 115
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 123 liqlerrtkyqssgimltfwlvalvcalaalrlfsetlhdnpopesasaslrlttlwi 182
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 116 SSLARVAHKKGELSMEDVWSLSKHESSDVNCRRLERLMQOELENFV-----GPDAA 165
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 183 tgl1-vrgyrgplsgdslnkedsqvyrvlwnkwkceaktkrpkykvysskda 241
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 166 -----SLRRVWICRRRLIIISYCLMTQLAGSGP 197
    :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Query Match 28.4%; Score 2075.5; DB 17; Length 1489;
 Best Local Similarity 34.0%; Pred. No. 3.9e-165;
 Matches 495; Conservative 266; Mismatches 497; Indels 199; Gaps 23;

```

Db 242 qpkesskvdaaneaveallvkspqkewmpslfkvlykftgpyflmsfffkahldmmfsgp 301
QY 198 AEMVHLELYQATSNQYSLILLVLGLLFTIVSRMSALTMALNRYRGVLRGAILTM 257
Db 302 q-llkllikftvdnpqwgqfyfvlvllfvtaclqtlvlhgyfhhcftvsgmktkaviga 360
QY 258 AFKIKLIKNIKIKERS--LGEELINICNSDQRMFEAAVGSLLAGRPVVALMINTVIL 315
Db 361 vyrtkavltinsrksstvgelvnmsvdqgrfmdlaelylmmsaprlqyljlyllwlnl 420
QY 316 GPTGLGSAVFLLEYFAMAFASRLATYFRRKCAATDERVQRMNVLATYIKPKMYAWK 375
Db 421 gpsvlagavmvlmrvnaamamcktyvahmkskdnrlkimeellngikvkllyawel 480
QY 376 AFSSQVQKIREERILKAKGFOSITVGVARIVVIVIASVTFESHMTLGEF--LTAQA 433
Db 481 arkdavlairegelvklksaylsavgtftwvcrpfvalcftavvltldemnlldagta 540
QY 434 FTVVTFNSMTFALKVTFPSYKSLSEASVAVDRFSLFMEEVH--MIKKNPASP---HI 488
Db 541 fvsialfhlrlfplnlpmvslsvgaesvlkrlflshelrpslerprvkdgsgtn 600
QY 489 KIEMKATLAMDSSHSIIONSPKLTPKMKKDKRASRCKEKYRQLQRTHEQAVALBQKH 548
Db 601 sltvnaftw----- 611
QY 549 LLLDDEPRSPPEEGKHHLHGLALQRTLSIDLEIOGKLVGICSGSGKTSLSIAI 608
Db 612 -----atsdpr-----tlngitfslpagaivavagvgcgklsllsal 649
QY 609 LGQMTLEGSIAISCTFPAVVAQOAMILNATLNDNLIFGKEYDEERYNSVLNCCRLPDLA 668
Db 650 laemkvqghvaiksgvayvpqgawlqndstrenllfqcgleepyrsvigcailpdl 709
QY 669 ILPSDLTEIGERGANLSGGQRQRTSLARALYSDBRSYIILDDPSALDAHVNHIENSAL 728
Db 710 lllpsgrtelqekyvlsgqkqrvslaraynsadilyfdpplsaavdahvghkhienvl 769
QY 729 --RKHLKSKTYFVTHOYLVDODEVIFMKKGCITRGTEHELMNLNDVATIF----- 781
Db 770 gpkgmllknktrllvchmsyLpqvdlivmsgklsengsygellardgaftaeftlryas 829
QY 782 -----NNLL-----GEPPEVINSKKEKTSQSKSODK 811
Db 830 teqegdaengvtysgyqgkkaekmgemlvtcdasgkqqrqlsssssgslsrhnm-- 887
QY 812 PKTGSIKKEKAVKPREGQVLDEKGGSVPSVYGVYIOAAGPLATLVIALPMLNNG 871
Db 888 -staellqkaekketwklmeadkaqtgkvklsvywdymkaiglflsfisif-lfmcnhv 945
QY 872 STAFSTWMLSWIKGSGNTTYTRGNETSYSDSMKDNPMHOVYASISALMSMVMILIKAI 931
Db 946 salasnyslvtldpbyngt---qehckvrlsvygalqsglavfgymsvstl----- 997
QY 932 RGVVEKSTLRASRLHDELFRRLRILSPMKFPDTPGTRILRFSKDMDEVDVRLPQAE 991
Db 998 -----gglasrclhndllhsilrpsmstfctcrpsgnlvrfrskeldvtvsmipevkl 1050
QY 992 MFIQWLVLFECVGMIAGVFWFLVAVCPVLVLSVLHVSVLI--RELKRLDNITQS 1048
Db 1051 mfmgsllfvngaclyllatpaaillppllgily--ffvgfvyassrqlkrlsvsrs 1107
QY 1049 PELSHTSSIOGLATIHAYNKGQELHRYQELLDNOAPFLFTCMARMLAVRLDLISIA 1108
Db 1108 pvyshfneelllvgvsvlrafeegeerfihgsdlkvdengkaypsivaanwlvavrlcevgnc 1167
QY 1109 LITTTGLMIVLHGQIPAYAGIAISYAVOLTGLOFTVRLASEFEARTSVERLNHXYK 1168
Db 1168 lylfaalavlsrhlslsglvglsysglqvttylnvlymssemctnvaerlkeyse 1227
QY 1169 TLTSEAPARIKKNKASPMWPOGEVTFENAEKRYRENTPLVLKVSFTIKPKKIGIYGR 1228
Db 1228 t-ekcapnglqgtrpsswpvgvgrvefrnyclyredldfvlrlnhvtlungekvglvgr 1286

```

```

QY 1229 TGSGKSIGMALFRLVELSGGCIKIDGVRISDGLADLRSKLSIIPQEPVLFSGTVRSNL 1288
Db 1287 tgaqksstlglfrlneasegeellidglniakglhndlrklllpqdpvlsgslrml 1346
QY 1289 DEFNQYTEDQIWDALEPRTHMKECIAQLPLKSEFVEMNGDNFSGEROLCTARALLRHC 1348
Db 1347 dfgsgsdeewvtslelahlkfvsalpkdkidhecaegenlsvgqrvclarallrkt 1406
QY 1349 KILIDETAMADTETDLLIOETREAFADCTMLTAHRLHTVLCSDRIMVLAQOVVEF 1408
Db 1407 kllyldeetaavldetddllqstlirtqfdeccvltlahrlntlmtytrvvlkgeigey 1466
QY 1409 DFPVSLNSDSSRFYAM 1425
Db 1467 gepsdl-l-qqrqlfysm 1482

RESULT 13
AAR54928
ID AAR54928 standard; Protein, 1531 AA.
XX
AC AAR54928;
XX
DT 14-OCT-1994 (first entry)
XX
DE Multidrug resistance protein.
XX
KW Multidrug resistance protein; MRP; H69AR; cancer cell line;
KM stem cell; cardiac muscle; transgenic animal.
OS Homo sapiens.
XX
PN W09410303-A.
XX
PD 11-MAY-1994.
XX
PF 27-OCT-1993; 93MO-CA00439.
XX
PR 27-OCT-1992; 92US-0966923.
XX
PR 08-MAR-1993; 93US-0029340.
XX
PA (TOOH ) UNIV QUEBENS KINGSTON.
XX
PI Cole SPC, Deeley RG.
XX
DR WPI: 1994-167460/20.
DR N-PSDB; AA065377.
XX
PT Multi-drug resistance gene - encodes protein capable of
PT conferring multi-drug resistance on cells, useful in diagnostic
PT and treatment methods
XX
PS Disclosure; Page 69-74; 101pp; English.
XX
CC The multidrug resistant cancer cell line H69AR (ATCC CRL 11350) was
CC used to identify cDNA encoding a novel protein associated with
CC multidrug resistance, MRP. MRP may be expressed in e.g.
CC hematopoietic stem cells or cardiac muscle, or in
CC transgenic animals, or can be used to raise antibodies.
XX
SQ Sequence 1531 AA;

Query Match 28.4%; Score 2075; DB 15; Length 1531;
Best Local Similarity 34.5%; Pred.No. 4.5e-165;
Matches 487; Conservative 261; Mismatches 481; Indels 184; Gaps 21.

QY 100 PYDNGGLSCMTFFSWLSLARIYAHKKGELSMEDVWSLSKHSDDVNCRLERLQOELENE 159
Db 209 pessasflsrllftwllgll-vrygrpqlsgsdwlnkedtsqvvvpylvkmwkkcak 267
QY 160 V-----GPDA-----SLRRVWTFRCRRLIL 181

```

```

Db 268 trkqpvkvyskdpapbkesskvdaneveallvkspqkempsltkvlytkfgyflm 327
QY 182 SIYCLMTOIAGSGPAPFAMVKHLLLEYQATENSQYSLILVGLGILLREIYSNGLATLWA 241
Db 328 sffikahndlmfmspgp-llklklfvyndtkapwgyfvtvllfvaciqtlvlyhfyh 386
QY 242 LMYRTGVRALGAILTMVAFKILKLNKIKES--LGELINICNDQGMFEAAVAGSLLAG 299
Db 387 lcfysgmrlkavlgayvkalvltnsarksvgelvnlmsvdaqfmdlatylnmlws 446
QY 300 GPVVALIGMIVNYIILCPGFLGSAVFIILPYPMAMPASRLTAIRKRCVATDERVOKM 359
Db 447 aplqvllalyllwlnlgsyvlaavmvlmvpvaamamtkcyqva hmkskdnrlk 506
QY 360 EVLYTIFIMYAMVAKFSSQVOKIREERRLLEKAGYFOSITGVAPIVVIVIASVYFS 419
Db 507 ellngllkvllaywelakfkdvajlrqgelvllkksaylsaavgtfvtwctprflvalcfa 566
QY 420 VHMUGFD--LTAQAFVTVVFNMSKTFALKVTPFSVKSLEASAVADVREKSLFMEVH 477
Db 567 vyvldemnlldaqatfvsalfnlfrplnlpmvlsivqaesvlkrlrlfshoele 626
QY 478 --MKNKPASB--HITKKNATLAWDSHSSIQNSPKLTPKKKKRKRSGKKEVQ 532
Db 627 pdsierprvkdggtnslcyvnaftw----- 653
QY 533 LQRTHQAVLAEGKHLLDSEDERPSEEGKHHLGLRLQRTLSIDLEIOEGLV 592
Db 654 -----arsdpp-----tlngltslpegalva 675
QY 593 ICGSVSGKTSLSAILGOMTLBGSIAISGFAYVAQAQMIATLADNLFEKEXDEE 652
Db 676 vvgvgvgcgkllslallaeamdkvegahvskgsavayvpqgawldnsirenlifgcjleap 735
QY 653 RYNSVLNCCCLRPALILPSSDLTEIGERGANISGQORQRLARALYSDMSIYLDPL 712
Db 736 yzrvldacalplpdllelpsgrtelgekynslsgqkqrvslaraynadllyldppl 795
QY 713 SALDAVGNHIFNSAI--RKHLKSKTVLFVTHQOLYVDCDEVIFMKEGCTTERGTHEEL 770
Db 796 savdahvghkhlfevnylpkgmlknktrilvshmsyypqvavllvmgsklsemgsygel 855
QY 771 MNLNGDYATIF-----NNLL-----GEPPEVEIN 795
Db 856 lardgataeflryastegeqdaengvtvysgpgkaakqmenqgmlyvdsagkqlgrjs 915
QY 796 SKRTSGSOKKSODKGPRTGSIRKKEKAVKPEGOLVLEEKGGSPVSVGYVIOAAG 855
Db 916 sssysgdishm---staelqkaakeetwklmeadkaqtgqvklsvywdymkaigl 972
QY 856 PLALVLMALFMLNVGSTAFSTWMLSYWIKOGSNTVTYRGNETSVSDMKDNPMOYYA 915
Db 973 flsfisflf-lfmcvhaalaenylslwtddplvngt---gehkvslsvygalisegl 1028
QY 916 SIYLSAAMVILKALNGVVFVKGTLRASSRLHDELFRRIILRSFPMKFPDTPGRIILNR 975
Db 1029 avfgymsavsl-----gglasrcbhvdlhslirpsmfferpsgnlvnrlf 1076
QY 976 SKMDDEVDRILPFOAEMFIQVILVFCVGMIAGVFPMFLAVGPLYLFLSVLHYSRVL 1035
Db 1077 skeldvdsampevllkmngslfnvlgacvillacpaaallppjlgily--ffvqrfy 1133
QY 1036 I---RELKRLDNIQSPFSLHITSIOGLATIHAYNKGQELFARQELLDNOAFLEFT 1092
Db 1134 vaasrqklriesverspyshfneltlvgvslrafegertlmgdtkvndqkaypsl 1193
QY 1093 CAMRLAVRDLISALITTTGLMIVLMHGOIPRAYGLAISAVOLTGLFQFVYRLASE 1152
Db 1194 vanlvavrllecycnclvlfafalvavlsrlhsaglvglsvsyglvctlylnwlvrmse 1253
QY 1153 TEARFTSVERNHITKTLLEAPARIKKNKAPSDMPQGEVTFENAEKREYENPLVLK 1212

```

```

Db 1254 metnivaverlkeyset-ekaaqpqigeltprsswpvgvrefnynclryredldfvlrh 1312
QY 1213 VSFTRKPKRIGIVGRYSGKSSLGMLFRLVLSGCCIRIDGVRISDGLADLRSLST 1272
Db 1313 lnvtingekvaylgrtgagksstlglfrfinesaegeellidglnaklgldhrlkll 1372
QY 1273 IPQEPVLFSGVRSNLDLPFNQYTTDOIMDALERTHMECIAQLPLKESSEVMGDNFSV 1332
Db 1373 lpgqpvlfsgslrmlnlpfsgysdeevwtslclahldkdvlsalpdklidhecaeggenlv 1432
QY 1333 GEROLGICARLIRHCKILILIDEATAAMDTELDLIOETIREAFADCTMLTIAHRLHTVL 1392
Db 1433 qgrqlvclatalrlklllyldeaavaadletdqlqstlrltgedctvltiahrlntlm 1492
QY 1393 GSDRIWVLAOGVPEPDPVSLNSDSSREFAM 1425
Db 1493 dytrvvlvldkgeiqegapell-qgrglfysm 1524

RESULT 14
AAR93153
ID AAR93153 standard; Protein: 1531 AA.
XX
AC AAR93153;
XX
XX 19-JUL-1996 (first entry)
XX
DE Multi-drug resistance protein.
XX
KW MRP; multi-drug resistance protein; small cell lung cancer;
KW protect; cytotoxicity; therapy; P-glycoprotein.
XX
XX Homo sapiens.
XX
FH Key
FT Region
FT /note= "putative membrane-spanning region"
FT 137..153
FT /note= "putative membrane-spanning region"
FT 175..191
FT /note= "putative membrane-spanning region"
FT 243..252
FT /note= "preferred immunogenic portion"
FT 365..381
FT /note= "putative membrane-spanning region"
FT 444..460
FT /note= "putative membrane-spanning region"
FT 466..482
FT /note= "putative membrane-spanning region"
FT 555..571
FT /note= "putative membrane-spanning region"
FT 591..607
FT /note= "putative membrane-spanning region"
FT 661..810
FT /note= "nucleotide binding fold"
FT 765..779
FT /note= "immunogenic portion"
FT 932..943
FT /note= "preferred immunogenic portion"
FT 969..985
FT /note= "putative membrane-spanning region"
FT 1028..1044
FT /note= "putative membrane-spanning region"
FT 1102..1118
FT /note= "putative membrane-spanning region"
FT 1205..1221
FT /note= "putative membrane-spanning region"
FT 1310..1469
FT /note= "nucleotide binding fold"
FT 1427..1441
FT /note= "immunogenic portion"
PN US5489519-A.

```


Tue Nov 13 11:55:00 2001

us-09-528-031-2.rag

Page 19

```

      |::||::|: || || ||::|
Db 1493 dytrvivi1dkgeiqeygapsdl1-qqrqlfysm 1524

```

Search completed: November 9, 2001, 17:45:56
Job time: 6854 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2001, 15:53:52 ; Search time 34.17 Seconds
(without alignments)
946.363 Million cell updates/sec

Title: US-09-528-031-2

Perfect score: 7308
Sequence: 1 MKDIDGKEIIIPSGYRSV.....DSSRFYAMFAAEKNAVKG 1437

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|------------------|
| 1 | 7308 | 100.0 | 1437 | 3 | US-09-061-400-2 |
| 2 | 7298 | 99.9 | 1453 | 2 | US-09-001-273-2 |
| 3 | 7298 | 99.9 | 1453 | 4 | US-08-843-459A-2 |
| 4 | 2082 | 28.5 | 1531 | 1 | US-08-463-092B-4 |
| 5 | 2082 | 28.5 | 1531 | 2 | US-08-462-109A-4 |
| 6 | 2082 | 28.5 | 1531 | 2 | US-08-460-907B-4 |
| 7 | 2082 | 28.5 | 1531 | 3 | US-08-463-179A-4 |
| 8 | 2082 | 28.5 | 1531 | 3 | US-08-461-384B-4 |
| 9 | 2075 | 28.4 | 1531 | 1 | US-08-141-893-2 |
| 10 | 2075 | 28.4 | 1531 | 1 | US-08-463-092B-2 |
| 11 | 2075 | 28.4 | 1531 | 2 | US-08-462-109A-2 |
| 12 | 2075 | 28.4 | 1531 | 2 | US-08-460-907B-2 |
| 13 | 2075 | 28.4 | 1531 | 3 | US-08-463-179A-2 |
| 14 | 2075 | 28.4 | 1531 | 3 | US-08-461-384B-2 |
| 15 | 2075 | 28.4 | 1531 | 3 | US-08-407-207A-2 |
| 16 | 2055.5 | 28.1 | 1528 | 1 | US-08-463-092B-6 |
| 17 | 2055.5 | 28.1 | 1528 | 2 | US-08-462-109A-6 |
| 18 | 2055.5 | 28.1 | 1528 | 2 | US-08-460-907B-6 |
| 19 | 2055.5 | 28.1 | 1528 | 3 | US-08-463-179A-6 |
| 20 | 2055.5 | 28.1 | 1528 | 3 | US-08-461-384B-6 |
| 21 | 2013.5 | 27.6 | 1622 | 4 | US-08-972-927-6 |
| 22 | 1991 | 27.2 | 1621 | 4 | US-08-972-927-3 |
| 23 | 1759 | 24.1 | 1581 | 4 | US-08-726-320-3 |
| 24 | 1759 | 24.1 | 1581 | 4 | US-09-208-716-3 |
| 25 | 1757.5 | 24.0 | 1580 | 4 | US-08-726-320-1 |
| 26 | 1757.5 | 24.0 | 1580 | 4 | US-09-208-716-1 |
| 27 | 1732 | 23.7 | 1581 | 4 | US-08-726-320-4 |

| | | | | | | |
|----|--------|------|------|---|-------------------|-------------------|
| 28 | 1732 | 23.7 | 1581 | 4 | US-09-208-716-4 | Sequence 4, Appl. |
| 29 | 1730 | 23.7 | 1477 | 3 | US-08-492-459-10 | Sequence 10, Appl |
| 30 | 1730 | 23.7 | 1477 | 3 | US-08-423-752-10 | Sequence 10, Appl |
| 31 | 1730 | 23.7 | 1477 | 3 | US-08-945-994-3 | Sequence 3, Appl |
| 32 | 1730 | 23.7 | 1477 | 4 | US-08-716-873-24 | Sequence 24, Appl |
| 33 | 1730 | 23.7 | 1477 | 4 | US-09-368-431-24 | Sequence 24, Appl |
| 34 | 1669.5 | 22.8 | 1581 | 2 | US-08-404-531B-6 | Sequence 6, Appl |
| 35 | 1669.5 | 22.8 | 1581 | 3 | US-08-476-900A-6 | Sequence 6, Appl |
| 36 | 1669.5 | 22.8 | 1581 | 3 | US-08-488-546A-6 | Sequence 6, Appl |
| 37 | 1669.5 | 22.8 | 1582 | 2 | US-08-404-531B-9 | Sequence 9, Appl |
| 38 | 1669.5 | 22.8 | 1582 | 3 | US-08-476-900A-9 | Sequence 9, Appl |
| 39 | 1669.5 | 22.8 | 1582 | 3 | US-08-488-546A-9 | Sequence 9, Appl |
| 40 | 1669.5 | 22.8 | 1582 | 4 | US-08-726-320-5 | Sequence 5, Appl |
| 41 | 1669.5 | 22.8 | 1582 | 4 | US-09-208-716-5 | Sequence 5, Appl |
| 42 | 1528.5 | 20.9 | 1498 | 2 | US-08-404-531B-28 | Sequence 28, Appl |
| 43 | 1528.5 | 20.9 | 1498 | 3 | US-08-476-900A-28 | Sequence 28, Appl |
| 44 | 1528.5 | 20.9 | 1498 | 3 | US-08-488-546A-28 | Sequence 28, Appl |
| 45 | 1513 | 20.7 | 1548 | 1 | US-08-463-092B-7 | Sequence 7, Appl |

ALIGNMENTS

```
RESULT 1
US-09-061-400-2
; Sequence 2, Application US/09061400
; Patent No. 6077936
; GENERAL INFORMATION:
; APPLICANT: SHYJAN, Andrew
; TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/061,400
; FILING DATE: 16-APRIL-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth A. Hanley
; REGISTRATION NUMBER: 33,505
; REFERENCE/DOCKET NUMBER: MNI-056CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-061-400-2

Query Match 100.0%, Score 7308, DB 3, Length 1437,
Best Local Similarity 100.0%, Pred. No. 0,
Matches 1437, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 MKDIDGKEIIIPSGYRSVREKTSSTGTHRDSDSKFRRTPLFCODALETAARAGIS 60
DB 1 MKDIDGKEIIIPSGYRSVREKTSSTGTHRDSDSKFRRTPLFCODALETAARAGIS 60
QY 61 LDASMSQRIIDDEHPKCKHYHGLSALKPIRTCTCKHQPVDNAGLFSGMTSWSLSLAR 120
```

|||||
Db LDASHSOLRLIDEBHPKGYHGLSALKPIRTTCKHOHPVDNAGLFSCMTFSWLSLAR 120
121 VAHKGELSMEDWMSLSKHESSDVNCRRLERLMOBELNEVPDASLRVYVWICRPLI 180
121 VAHKGELSMEDWMSLSKHESSDVNCRRLERLMOBELNEVPDASLRVYVWICRPLI 180
121 VAHKGELSMEDWMSLSKHESSDVNCRRLERLMOBELNEVPDASLRVYVWICRPLI 180
181 LSVICMLITQAGSGPAFWKHLETTQATESNLQYSLLVGLLLEIYRSMSLATW 240
181 LSVICMLITQAGSGPAFWKHLETTQATESNLQYSLLVGLLLEIYRSMSLATW 240
181 LSVICMLITQAGSGPAFWKHLETTQATESNLQYSLLVGLLLEIYRSMSLATW 240
181 LSVICMLITQAGSGPAFWKHLETTQATESNLQYSLLVGLLLEIYRSMSLATW 240
241 ALNRTGVRRLGALITMAFKILKLNKEKSLGELINICNDGQRMFEAAVSLAGG 300
241 ALNRTGVRRLGALITMAFKILKLNKEKSLGELINICNDGQRMFEAAVSLAGG 300
241 ALNRTGVRRLGALITMAFKILKLNKEKSLGELINICNDGQRMFEAAVSLAGG 300
301 PVVALIGITVAVIILGPGFISAVFIIFPAMMFASRLTAYFRKCAANDEROKANE 360
301 PVVALIGITVAVIILGPGFISAVFIIFPAMMFASRLTAYFRKCAANDEROKANE 360
301 PVVALIGITVAVIILGPGFISAVFIIFPAMMFASRLTAYFRKCAANDEROKANE 360
301 PVVALIGITVAVIILGPGFISAVFIIFPAMMFASRLTAYFRKCAANDEROKANE 360
361 VLTYIKFKITMAWYKAFSOSVOKITREERRLERKAGYFQSIITVGAPITVVYIASVTFPSV 420
361 VLTYIKFKITMAWYKAFSOSVOKITREERRLERKAGYFQSIITVGAPITVVYIASVTFPSV 420
361 VLTYIKFKITMAWYKAFSOSVOKITREERRLERKAGYFQSIITVGAPITVVYIASVTFPSV 420
361 VLTYIKFKITMAWYKAFSOSVOKITREERRLERKAGYFQSIITVGAPITVVYIASVTFPSV 420
421 HMTIGFDLTAQAATVTVTVENSMTFALKVTPFSYKSLSEASVANDREKSLMEVYHAK 480
421 HMTIGFDLTAQAATVTVTVENSMTFALKVTPFSYKSLSEASVANDREKSLMEVYHAK 480
421 HMTIGFDLTAQAATVTVTVENSMTFALKVTPFSYKSLSEASVANDREKSLMEVYHAK 480
421 HMTIGFDLTAQAATVTVTVENSMTFALKVTPFSYKSLSEASVANDREKSLMEVYHAK 480
481 NKPAIPHIKEMKATLAMDSSHSIONSPLKTPKMKDKRASRGKREKREBOLRTEHOA 540
481 NKPAIPHIKEMKATLAMDSSHSIONSPLKTPKMKDKRASRGKREKREBOLRTEHOA 540
481 NKPAIPHIKEMKATLAMDSSHSIONSPLKTPKMKDKRASRGKREKREBOLRTEHOA 540
481 NKPAIPHIKEMKATLAMDSSHSIONSPLKTPKMKDKRASRGKREKREBOLRTEHOA 540
541 VLABQKHLIDSDERBSPBEEBEGKHILGLRLQRLHSLIDLEIOEGKLVGICGSVSG 600
541 VLABQKHLIDSDERBSPBEEBEGKHILGLRLQRLHSLIDLEIOEGKLVGICGSVSG 600
541 VLABQKHLIDSDERBSPBEEBEGKHILGLRLQRLHSLIDLEIOEGKLVGICGSVSG 600
541 VLABQKHLIDSDERBSPBEEBEGKHILGLRLQRLHSLIDLEIOEGKLVGICGSVSG 600
601 KTSLSIALIGOMTLEGGISAIISGTPAYAAQAMILNATLRNIIIFGKRYDERNSVANS 660
601 KTSLSIALIGOMTLEGGISAIISGTPAYAAQAMILNATLRNIIIFGKRYDERNSVANS 660
601 KTSLSIALIGOMTLEGGISAIISGTPAYAAQAMILNATLRNIIIFGKRYDERNSVANS 660
601 KTSLSIALIGOMTLEGGISAIISGTPAYAAQAMILNATLRNIIIFGKRYDERNSVANS 660
661 CCLRPDLAIPSSDLTEIGERGANLSSGQRORIISLARALYSDRSIYILDDPLSALDAHVG 720
661 CCLRPDLAIPSSDLTEIGERGANLSSGQRORIISLARALYSDRSIYILDDPLSALDAHVG 720
661 CCLRPDLAIPSSDLTEIGERGANLSSGQRORIISLARALYSDRSIYILDDPLSALDAHVG 720
661 CCLRPDLAIPSSDLTEIGERGANLSSGQRORIISLARALYSDRSIYILDDPLSALDAHVG 720
721 NHINSAIRKHLKSTVLFVTHOLOYLVDCDEVIFPKEGCTTERGTHEMLNMGDAFTI 780
721 NHINSAIRKHLKSTVLFVTHOLOYLVDCDEVIFPKEGCTTERGTHEMLNMGDAFTI 780
721 NHINSAIRKHLKSTVLFVTHOLOYLVDCDEVIFPKEGCTTERGTHEMLNMGDAFTI 780
721 NHINSAIRKHLKSTVLFVTHOLOYLVDCDEVIFPKEGCTTERGTHEMLNMGDAFTI 780
781 FNNLLGFTPEVEINSKKEKTSKSGOKSODKPKTSGIKKEKAVKPEEGOLVLEKKGGS 840
781 FNNLLGFTPEVEINSKKEKTSKSGOKSODKPKTSGIKKEKAVKPEEGOLVLEKKGGS 840
781 FNNLLGFTPEVEINSKKEKTSKSGOKSODKPKTSGIKKEKAVKPEEGOLVLEKKGGS 840
781 FNNLLGFTPEVEINSKKEKTSKSGOKSODKPKTSGIKKEKAVKPEEGOLVLEKKGGS 840
841 VPMVYGYVYIOAAGGPIALFVLMALFMLNVGSAFSTWMLYWKSGGNTYVRNGNTS 900
841 VPMVYGYVYIOAAGGPIALFVLMALFMLNVGSAFSTWMLYWKSGGNTYVRNGNTS 900
841 VPMVYGYVYIOAAGGPIALFVLMALFMLNVGSAFSTWMLYWKSGGNTYVRNGNTS 900
841 VPMVYGYVYIOAAGGPIALFVLMALFMLNVGSAFSTWMLYWKSGGNTYVRNGNTS 900
901 VDSKMDKPNHMOYATYALSMAVWLILKAIKGVYFVKGTRASSRLDELFRILRISPM 960
901 VDSKMDKPNHMOYATYALSMAVWLILKAIKGVYFVKGTRASSRLDELFRILRISPM 960
901 VDSKMDKPNHMOYATYALSMAVWLILKAIKGVYFVKGTRASSRLDELFRILRISPM 960
901 VDSKMDKPNHMOYATYALSMAVWLILKAIKGVYFVKGTRASSRLDELFRILRISPM 960
961 KEFDTTPGRLINPFSKMDMDVRLPFOAEMFIONVILVFCVGMINGVPMFLVAVGP 1020
961 KEFDTTPGRLINPFSKMDMDVRLPFOAEMFIONVILVFCVGMINGVPMFLVAVGP 1020
961 KEFDTTPGRLINPFSKMDMDVRLPFOAEMFIONVILVFCVGMINGVPMFLVAVGP 1020
961 KEFDTTPGRLINPFSKMDMDVRLPFOAEMFIONVILVFCVGMINGVPMFLVAVGP 1020
1021 LVILFSLVLIHSRVILRLKRLDNTOSPLSHITSSIOGLATTHAYKGOEFLHRYOEL 1080
1021 LVILFSLVLIHSRVILRLKRLDNTOSPLSHITSSIOGLATTHAYKGOEFLHRYOEL 1080
1021 LVILFSLVLIHSRVILRLKRLDNTOSPLSHITSSIOGLATTHAYKGOEFLHRYOEL 1080
1021 LVILFSLVLIHSRVILRLKRLDNTOSPLSHITSSIOGLATTHAYKGOEFLHRYOEL 1080
1081 LDDNOAPPELTCAMRWLAVALDLISALITTTGLMIVLMHGOLPAPVAGLAISYAVOLT 1140
1081 LDDNOAPPELTCAMRWLAVALDLISALITTTGLMIVLMHGOLPAPVAGLAISYAVOLT 1140
1081 LDDNOAPPELTCAMRWLAVALDLISALITTTGLMIVLMHGOLPAPVAGLAISYAVOLT 1140
1081 LDDNOAPPELTCAMRWLAVALDLISALITTTGLMIVLMHGOLPAPVAGLAISYAVOLT 1140
1141 GLFOFTVRLASETARTSVIRINHYITLSLEAPARIKKNAPSDMFOBEVPEENEM 1200
1141 GLFOFTVRLASETARTSVIRINHYITLSLEAPARIKKNAPSDMFOBEVPEENEM 1200
1141 GLFOFTVRLASETARTSVIRINHYITLSLEAPARIKKNAPSDMFOBEVPEENEM 1200

Db 1141 GLFOFTVRLASETARTSVIRINHYITLSLEAPARIKKNAPSDMFOBEVPEENEM 1200
Qy 1201 RYRENPLVLKAVFTTIPKPKKIGVGTGSGKSLGMALEFLVELSGGCIKIDGVRI 1260
Db 1201 RYRENPLVLKAVFTTIPKPKKIGVGTGSGKSLGMALEFLVELSGGCIKIDGVRI 1260
Qy 1261 IGLADSKSLIIOEPVLFSGYVRSNIDPPNOYTEDQIMVALERTHKECIAOLPLKLE 1320
Db 1261 IGLADSKSLIIOEPVLFSGYVRSNIDPPNOYTEDQIMVALERTHKECIAOLPLKLE 1320
Qy 1321 SEVENGNFSVGEKOLICIRALLRHCKILILDEPATAMTERDILLIOETIRAPADCT 1380
Db 1321 SEVENGNFSVGEKOLICIRALLRHCKILILDEPATAMTERDILLIOETIRAPADCT 1380
Qy 1381 MLTIAHRLHTVLGSDRIWVLAQGVVEEDTSPVLLSNDSSRFYMAFAENKVAVK 1437
Db 1381 MLTIAHRLHTVLGSDRIWVLAQGVVEEDTSPVLLSNDSSRFYMAFAENKVAVK 1437
RESULT 2
US-09-001-273-2
: Sequence 2, Application US/09001273
: Patent No. 5994130
: GENERAL INFORMATION:
: APPLICANT: SHIVAN, Andrew
: TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
: TITLE OF INVENTION: POLYPEPTIDE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Testa, Hurwitz & Thibault
: STREET: 125 High St.
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/001,273
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: FENTON, Gillian M
: REGISTRATION NUMBER: 36,508
: REFERENCE/DOCKET NUMBER: MIL-001
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1453 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-001-273-2
Query Match 99.9% Score 7298: DB 2: Length 1453;
Best Local Similarity 99.9% Pred. No. 0:
Matches 1436: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
Qy 1 MKDIDIGKEYIIPSPGYSVREERTSGTHDRDREDSKFRTRPPLCOCALETAARAEGLS 60
Db 17 MKDIDIGKEYIIPSPGYSVREERTSGTHDRDREDSKFRTRPPLCOCALETAARAEGLS 76
Qy 61 LDASHSOLRLIDEBHPKGYHGLSALKPIRTTCKHOHPVDNAGLFSCMTFSWLSLAR 120
Db 77 LDASHSOLRLIDEBHPKGYHGLSALKPIRTTCKHOHPVDNAGLFSCMTFSWLSLAR 136
Qy 121 VAHKGELSMEDWMSLSKHESSDVNCRRLERLMOBELNEVPDASLRVYVWICRPLI 180

```
Db 137 VAHKGELSMEDVWSLSHRESSDVNCRLERLMQBELNEVGDAASLRVWVIFCRTL 196
Qy 181 LSIYVCLMTOLAGSGPAPMKHLELYQATESNLQYSLLVGLLLEIYVWSGLATW 240
Db 197 LSIYVCLMTOLAGSGPAPMKHLELYQATESNLQYSLLVGLLLEIYVWSGLATW 256
Qy 241 ALNVTGRLGAILTMAFKLKLKNIKESLGLINICSDGORMEPAAVSLLAGG 300
Db 257 ALNVTGRLGAILTMAFKLKLKNIKESLGLINICSDGORMEPAAVSLLAGG 316
Qy 301 PVVALIGMIYVNIILGPTGLGSAVFIIFYPAMPASRLTAFFRRKCAVADEROKME 360
Db 317 PVVALIGMIYVNIILGPTGLGSAVFIIFYPAMPASRLTAFFRRKCAVADEROKME 376
Qy 361 VLTITKFKMTAWAKAFSOSVOKTREERRLIEKAGYQSTIVGAPYVVIASVVTESV 420
Db 377 VLTITKFKMTAWAKAFSOSVOKTREERRLIEKAGYQSTIVGAPYVVIASVVTESV 436
Qy 421 HMTIGFDLTAOAFTVVPVNSMFPALKVTPTPSVKSLSASAVADRFKSLFMEEVHMK 480
Db 437 HMTIGFDLTAOAFTVVPVNSMFPALKVTPTPSVKSLSASAVADRFKSLFMEEVHMK 496
Qy 481 NKPSAPRIKIEKNATLAMDSHSSSIONSPKLTPKMKDKRASRGKKEKVRQLQRTHEQA 540
Db 497 NKPSAPRIKIEKNATLAMDSHSSSIONSPKLTPKMKDKRASRGKKEKVRQLQRTHEQA 556
Qy 541 VLABQKGLLLDSRPSPEBEKHHGLRLQRTLSIDLEIOEGKLVICGSVSG 600
Db 557 VLABQKGLLLDSRPSPEBEKHHGLRLQRTLSIDLEIOEGKLVICGSVSG 616
Qy 601 KTSLSALIGOMTLEGGISATSGTAPYAAOAMINATLRDNLIEGKEDERYSVNS 660
Db 617 KTSLSALIGOMTLEGGISATSGTAPYAAOAMINATLRDNLIEGKEDERYSVNS 676
Qy 661 CCLRPDLALIPSSDLTEIGERGANISGGORORISLARALYSDRSIYILDDPLSALDAHVG 720
Db 677 CCLRPDLALIPSSDLTEIGERGANISGGORORISLARALYSDRSIYILDDPLSALDAHVG 736
Qy 721 NHIFSAIRKHLKSTVYLFVTHOLOLVDCDEVIPKESGCTERGTHERLMLNDYATI 780
Db 737 NHIFSAIRKHLKSTVYLFVTHOLOLVDCDEVIPKESGCTERGTHERLMLNDYATI 796
Qy 781 FNNLLGTPPEINRSKETSODKPGKTSIKKEKAVKEEGOLVLEKGGOS 840
Db 797 FNNLLGTPPEINRSKETSODKPGKTSIKKEKAVKEEGOLVLEKGGOS 856
Qy 841 VPMVSYGYVYIOAAGPLAFVYLMALFMLNVGSTAFSTWMLSYWIKOGSGNTVTGNETS 900
Db 857 VPMVSYGYVYIOAAGPLAFVYLMALFMLNVGSTAFSTWMLSYWIKOGSGNTVTGNETS 916
Qy 901 VDSKDKNPHMOYATISALSMVAWMLIKAIRGVYFKGLTRASSRLHDELFRRLRSPM 960
Db 917 VDSKDKNPHMOYATISALSMVAWMLIKAIRGVYFKGLTRASSRLHDELFRRLRSPM 976
Qy 961 KEFDTTPGRIILNRSKMDDEVNRLPFOAEMFIONVILVFCVGMINGVFMFLVAVGP 1020
Db 977 KEFDTTPGRIILNRSKMDDEVNRLPFOAEMFIONVILVFCVGMINGVFMFLVAVGP 1036
Qy 1021 LVILFVSLHIVSRVILRELKRLDNTGSPFLSHITSSIOGLATIHAYNKGQEFRLRYOEL 1080
Db 1037 LVILFVSLHIVSRVILRELKRLDNTGSPFLSHITSSIOGLATIHAYNKGQEFRLRYOEL 1096
Qy 1081 LDDNAPPELFECARMLAVRLDLISALITTTGMLIVLMHQIPRAYAGLAISAVOLT 1140
Db 1097 LDDNAPPELFECARMLAVRLDLISALITTTGMLIVLMHQIPRAYAGLAISAVOLT 1156
Qy 1141 GLFOFTVRLASFTEARFTSVERINHYITLSLEAPARIKNKAPSPDMQEGEVTFENAM 1200
Db 1157 GLFOFTVRLASFTEARFTSVERINHYITLSLEAPARIKNKAPSPDMQEGEVTFENAM 1216
Qy 1201 RYREMLPLVKKVSPFTIKPEKIGIVGTGSGKSSIGMALFRLVELSGGCIRKIDGVRI 1260
Db 1260 RYREMLPLVKKVSPFTIKPEKIGIVGTGSGKSSIGMALFRLVELSGGCIRKIDGVRI 1260
```

```
Db 1217 RYREMLPLVKKVSPFTIKPEKIGIVGTGSGKSSIGMALFRLVELSGGCIRKIDGVRI 1276
Qy 1261 IGLADLRKSLSTIPPEPVLFSTVSNNDPFWQYTEDQIMDLERTHKECIAOLPLKE 1320
Db 1277 IGLADLRKSLSTIPPEPVLFSTVSNNDPFWQYTEDQIMDLERTHKECIAOLPLKE 1336
Qy 1321 SEVMNGDNFSVGEQOLICARALLRHCKIILDEATAMDETDLLOETIREFADCT 1380
Db 1337 SEVMNGDNFSVGEQOLICARALLRHCKIILDEATAMDETDLLOETIREFADCT 1396
Qy 1381 MLTIAHRLHTVLSDRINWYLAQGVVEFTPSVLLSNDSSRFYAMFAAENKVAVG 1437
Db 1397 MLTIAHRLHTVLSDRINWYLAQGVVEFTPSVLLSNDSSRFYAMFAAENKVAVG 1453

RESULT 3
US-08-843-459A-2
: Sequence 2, Application US/08843459A
: Patent No. 6162616
: GENERAL INFORMATION:
: APPLICANT: SHYUN, Andrew
: TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
: TITLE OF INVENTION: POLYPEPTIDE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/843.459A
: FILING DATE: 16-Apr-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanley, Elizabeth A.
: REGISTRATION NUMBER: 33,505
: REFERENCE/DOCKET NUMBER: MNI-056 (formerly MIL-001)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)742-4214
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1453 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-843-459A-2

Query Match 99.9% Score 7298: DB 4: Length 1453:
Best Local Similarity 99.9%: Pred. No. 0:
Matches 1436: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

Qy 1 MKDIDIGKEYIIPSPGYSVRERTSGTHRDREDSKFRRTPLBCODALETAARAEGLS 60
Db 17 MKDIDIGKEYIIPSPGYSVRERTSGTHRDREDSKFRRTPLBCODALETAARAEGLS 76
Qy 61 LDASHMSQRLIIDEHPKGYKHGLSALKPIRTTCKHQHPVDNAGLFSCMTFSWLSLAR 120
Db 77 LDASHMSQRLIIDEHPKGYKHGLSALKPIRTTCKHQHPVDNAGLFSCMTFSWLSLAR 136
Qy 121 VAHKGELSMEDVWSLSHRESSDVNCRLERLMQBELNEVGDAASLRVWVIFCRTL 180
Db 137 VAHKGELSMEDVWSLSHRESSDVNCRLERLMQBELNEVGDAASLRVWVIFCRTL 196
Qy 181 LSIYVCLMTOLAGSGPAPMKHLELYQATESNLQYSLLVGLLLEIYVWSGLATW 240
Db 240 LSIYVCLMTOLAGSGPAPMKHLELYQATESNLQYSLLVGLLLEIYVWSGLATW 240
```

|||||
Db 197 LSTVCLMTQAGSGAPFAMKHLLETQATSNLQYSLLVLELLLETVRSNLSLW 256
QY 241 ALNRTGVRGALITTAFAFKILKLNKEKSLGELLINCSNDQGRFEAAVSLAG 300
Db 257 ALNRTGVRGALITTAFAFKILKLNKEKSLGELLINCSNDQGRFEAAVSLAG 316
QY 301 PVVAITGMIVNIIIGPTGGLGSAVFLFETPAMFASRLTAFFRRKCAVATDERVQKME 360
Db 317 PVVAITGMIVNIIIGPTGGLGSAVFLFETPAMFASRLTAFFRRKCAVATDERVQKME 376
QY 361 VLVYIKFKIKNWAWKAFSOSQKIREERRLIEKAGFOSITVAVPIVVAIVASVTFSV 420
Db 377 VLVYIKFKIKNWAWKAFSOSQKIREERRLIEKAGFOSITVAVPIVVAIVASVTFSV 436
QY 421 HMTLGFDTAAQAFVTVTFSMTFALKVTFYSKSLSEASVADRFKSLTMEEVHMK 480
Db 437 HMTLGFDTAAQAFVTVTFSMTFALKVTFYSKSLSEASVADRFKSLTMEEVHMK 496
QY 481 NKPASPIKIEKKNATLAWDSHSSSIONSPKLPKMKDKRASGKKEKYAROLQTEHOA 540
Db 497 NKPASPIKIEKKNATLAWDSHSSSIONSPKLPKMKDKRASGKKEKYAROLQTEHOA 556
QY 541 VLAQKCHLLDSEERSPEEEKHHLGHLRLQRTLHSDLEIOEGKLVIGTGSVSG 600
Db 557 VLAQKCHLLDSEERSPEEEKHHLGHLRLQRTLHSDLEIOEGKLVIGTGSVSG 616
QY 601 KTSLSAIILOMTLLECSIASGTFAYAAOAWILNATLRDNIIFGKEDYDERNSVLNS 660
Db 617 KTSLSAIILOMTLLECSIASGTFAYAAOAWILNATLRDNIIFGKEDYDERNSVLNS 676
QY 661 CCLRPDLAIISSDLEIGERGANLGGQORISLARALYSDRSIIYLLDPLSALDAHV 720
Db 677 CCLRPDLAIISSDLEIGERGANLGGQORISLARALYSDRSIIYLLDPLSALDAHV 736
QY 721 NHFNSAIRKHKSKYLVFTHQOYLVDCEVLFMEKGGCTTEGTHHEMLNMGDAFI 780
Db 737 NHFNSAIRKHKSKYLVFTHQOYLVDCEVLFMEKGGCTTEGTHHEMLNMGDAFI 796
QY 781 FNNLLGETPPEVEINSKKTSGSOKKSQDKPKTSGIKKEKAVPREGQOLVLEKSGGS 840
Db 797 FNNLLGETPPEVEINSKKTSGSOKKSQDKPKTSGIKKEKAVPREGQOLVLEKSGGS 856
QY 841 VPMSVYGVYIOAGGPIAFVLYALFMLNNGSTAFSTWMLSYTKKSGSGNTTVRGNETS 900
Db 857 VPMSVYGVYIOAGGPIAFVLYALFMLNNGSTAFSTWMLSYTKKSGSGNTTVRGNETS 916
QY 901 VSDSMKDNPHQOYASISYALSMAYMLILKAIKRGVYFVKGTIRASSRLHDELFRILNSPM 960
Db 917 VSDSMKDNPHQOYASISYALSMAYMLILKAIKRGVYFVKGTIRASSRLHDELFRILNSPM 976
QY 961 KFPDPTPTGRIINFRSKDMDEVDVRLPFOAEMFIQNYLVFECVGMILAGVPMFLVAVGP 1020
Db 977 KFPDPTPTGRIINFRSKDMDEVDVRLPFOAEMFIQNYLVFECVGMILAGVPMFLVAVGP 1036
QY 1021 LVILFSVLHIVSRKLEKRLDNTOSPFSLSHITSSIOGLATTHAYNKGQEFHRYOEL 1080
Db 1037 LVILFSVLHIVSRKLEKRLDNTOSPFSLSHITSSIOGLATTHAYNKGQEFHRYOEL 1096
QY 1081 LDDQAQFPFLPTCAMRLAVRLDLSIALITTTGIMLVLMHGOIPPAVAGAIISYAVOLT 1140
Db 1097 LDDQAQFPFLPTCAMRLAVRLDLSIALITTTGIMLVLMHGOIPPAVAGAIISYAVOLT 1156
QY 1141 GLFOFTVRLASETEARTSYERINHYIKTISLEAPARIKKNAPSPDPOEGEVTFENAE 1200
Db 1157 GLFOFTVRLASETEARTSYERINHYIKTISLEAPARIKKNAPSPDPOEGEVTFENAE 1216
QY 1201 RYRENLPVLVAKKYSFTTKPKKIGIVGTSGKSSIGMALFRLVELSGCCIKIDGVARS 1260
Db 1217 RYRENLPVLVAKKYSFTTKPKKIGIVGTSGKSSIGMALFRLVELSGCCIKIDGVARS 1276
QY 1261 IGLADLSKSLIIQOEVLVSGTVRSNLDPPNOYTEDOINDALERTHMKECIAOLPIKLE 1320
|||||

Db 1277 IGLADLSKSLIIQOEVLVSGTVRSNLDPPNOYTEDOINDALERTHMKECIAOLPIKLE 1336
QY 1321 SEVENDGNFVSGEROLLICARALLRCKIILIDEATPAADTETDLLIOETIREAFADCT 1380
Db 1337 SEVENDGNFVSGEROLLICARALLRCKIILIDEATPAADTETDLLIOETIREAFADCT 1396
QY 1381 MLTIAHRLHYLGSDRIMVLAQGOVEFDPSVLLSNDSSRFYAMFAAEKKAHVKG 1437
Db 1397 MLTIAHRLHYLGSDRIMVLAQGOVEFDPSVLLSNDSSRFYAMFAAEKKAHVKG 1453
|||||

RESULT 4
US-08-463-092B-4
Sequence 4, Application US/08463092B
Patent No. 5766880
GENERAL INFORMATION:
APPLICANT: Cole, Deeley, Roger G.
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PARTEO RESEARCH & DEVELOPMENT INNOVATIONS
STREET: Queen's University at Kingston
CITY: Kingston
STATE: Ontario
COUNTRY: CANADA
ZIP: K7L 3N6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,092B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: Q1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-6853
TELEFAX: (613) 545-2342
INFORMATION FOR SRO ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1531 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-092B-4
Query Match 28.5%; Score 2082; DB 1; Length 1531;
Best Local Similarity 34.5%; Pred. No. 1.8e-195;
Matches 488; Conservative 262; Mismatches 479; Indels 184; Gaps 21;

| | | | |
|----|------|---|------|
| Qy | 100 | PVAGLESCMTSLSSSLARVHKHKGELSMEDWSLSKHSBDVNCNRBRLEHQEOLNE | 155 |
| Db | 209 | PESSAKLSRTTWMVITGLI-VAGYRQPLESGDLSLNEKDETEQVYVPLVKNMKKECAK | 267 |
| Qy | 160 | V-----GPDA-----SLRWVJFCRTRIL | 181 |
| Db | 268 | TRKRPVWVYSSKDPAPQKRESSVPDANEVEVLLYKSPQKKNPNSLFLVYIKTGPPIYLM | 327 |
| Qy | 182 | SIYCLMTLOLAFSGPAPFVWKHLELYTOATESNLQYSLLVIGLLLTETVRSWLSLTAWA | 241 |
| Db | 328 | SEFFKALHIDLMFESGPO-ILKLIFKPVNDTKAPDMQGYEYVLLLEPVACLOTLVLYHQFH | 386 |
| Qy | 242 | LNRYTGRLAGALLTMAFKKILKLKNIKRS--LGELINICSDQGRPEAANSGSLLAG | 299 |
| Db | 387 | ICFSGKRITAYAGVYRRKALVTFSARKSSVGEIIVMLSVDAQRFMDATYINMTWS | 446 |
| Qy | 300 | GPPVAILGIMTYNVLIGPFGLSAVFLLFYAPMFPASLTAYEPFKCAVATDEYQCKN | 359 |
| Db | 447 | APLOYILALVLLMLNIGPVSILVAVAMVLMFPVNAVMAKTITYVAAHMSKDNRIKLMN | 506 |
| Qy | 360 | EVLIVYIKFIMYAMVAFSOSYOKIEEERILIERAGYQOSITTVAVIIVVIASVVS | 419 |
| Db | 507 | EILNGIVALKYAMELAFKOKVLAIQEELKYLKKSAYLSANAGTFWVCTPRIVALCFPA | 566 |
| Qy | 420 | VHMTLGD--LTAQAFTVYVFNSTFALKYTPPSVKSLSASVAVDPRFSLFMEVBH | 477 |
| Db | 567 | VYVVIDENNITLDQTAFAVSIALTEINILRPLNIIPLWISSIYASVILKRLRFLSHELE | 626 |
| Qy | 478 | --MKKNPASP---HIKIEKNMTLAMDSSHSSIONSPLTPFKMKKDKRASGKKEKRYQ | 532 |
| Db | 627 | POSTERRPVKDDGGSTSVYKNAFTFW----- | 653 |
| Qy | 533 | LQRTHOAVLEAQKGHLLDSDRPSPEEGBKHHLGLRLQRTLHSDILETOEGKLVG | 592 |
| Db | 654 | -----ARSDP-----TLNGTTSIRPGALYA | 675 |
| Qy | 593 | ICGSVSGGKTSLSAILGQMTLLEGSIAISGTPRAYVAQOAMLTNRDNIIFGEKDEE | 652 |
| Db | 676 | VVGVGCGGKSSLLSALIAEWDKVEGHVAIKGSVAAYVPOQAMQNDISRENILFGCOLEBP | 735 |
| Qy | 653 | RYNVLVNSCCLRPDLILPSSDLTEIGEGANVSGGORISLAAVATSDRSITYLDPL | 712 |
| Db | 736 | YKRSYVIOACALLPDLLETPLEBSGDTETIGEGVNLSGGQKORVSLARAVYSNADITYLDP | 795 |
| Qy | 713 | SALDAVGHNIFFNSAI--RKHLKSTVLEFVTHOLOLYDQCDVIFPMKBCGCTERTGHEEL | 770 |
| Db | 796 | SAVDAHGKHIFENVIGPKMLKNKRIILTYHSMSTLPQVDVLIYMSGKSIEMKSYOEL | 855 |
| Qy | 771 | MNLNGDYATTF-----NNLL-----GETPVEIN | 795 |
| Db | 856 | LARGGAFAEFLRYASTAEODAEENGVTGVSQPKCAKOMENGMLYTDISAGKOLQRLS | 915 |
| Qy | 796 | SKRTSOSQKKSDQKGRKYSIKKERAUVRPEBOLVQLEEKQSVPMYSVYVYIOAGG | 855 |
| Db | 916 | SSSYSDSDISHHN---STWELQAKAAKKEETWKLMEAKQOTGYVKSLSYWDYKAIAGL | 972 |
| Qy | 856 | PLAFLVLMALFMLNVGSTASTWMLSUYMIKQSGGNTVYRGNETSVDSPKDNPMQYYA | 915 |
| Db | 973 | FISLSTLF-LPMCNHVSALASNWLSLMTDPIVNT---QEHTRYKLSYVGCALGISOCI | 1028 |
| Qy | 916 | SIYLSAAMVMLIKAIRGVVFKGLRASSRLHDELFRILRSPKFPDTPGTGILNRF | 975 |
| Db | 1029 | AVEFYSAUVSI-----GGILASRCLHYDLHLSILRSPSFEFRPNSGMLVNRF | 1076 |
| Qy | 976 | SKMDDEVDRLPQOAEFIONVTLVFCVGMLAGVPPMLVAVGVLVILFSLVLIHVSVL | 1035 |
| Db | 1077 | SKELDVTDSMPEYIKFEMKSLNVLNIGACTIVILATPIAIIITPPLGLIY--FVQGFY | 1133 |
| Qy | 1036 | I---RELKRDNIOTSPFLSHITSIOGLATIHAYKKQGEFLHRYOELDDQAFFLFT | 1092 |
| Db | 1134 | VASSRQJLRKLESRSRVSYVSHFNETILGSVIRAFEEQGRFLHQDGLKVDENQKRYYSI | 1193 |
| Qy | 1093 | CAMRWLAVRDLISIALITTTGMLIVLHMQIPRAYAGLAISYAVOLGELQFQTVRLASE | 1152 |

| Query Match | 28.5% | Score 2082 | DB 2 | Length 1531 |
|-----------------------|------------------|--|------------|-------------|
| Best Local Similarity | 34.5% | Pred. No. 1.8e-195 | | |
| Matches 488 | Conservative 262 | Mismatch 479 | Indels 184 | Gaps 21 |
| QY | 100 | PVDNAGLPSCHTFMFWLSSILARVAHKKGETSMEDVSLSKSHESDVNCRRLERLMOEELINE | 159 | |
| Db | 209 | PSSASFSLRITFWWITGLI--VGRGRÖPLEGSDLMSLNKEDISBDQVPLVYANKMECK | 267 | |
| QY | 160 | V-----GPDAA----- | 181 | |
| Db | 268 | TRKÖPKVYVYSSKDAÖPKRESSKVDANEVEALIVKSPÖKENPSSLFKYLYTFPGPYFLM | 327 | |
| QY | 182 | STVCLMTIÖLAGSGPAPFVKKHLEETOATESNLOXSLLVGLITELIVRSMSLATAWA | 241 | |
| Db | 328 | SPEEFKAHIDLMFSGPÖ--ILKLLIKFVNDTKAPMÖGFTYVLLLEVTACTÖLVLVHÖYH | 386 | |
| QY | 242 | LNVRGTVRLRGAILTMARFKKILIKLNIEKS--IGELINICSDGÖRMEEAAGVSLLAG | 299 | |
| Db | 387 | ICFVSGMRKIKTEVAVYKRVKALVITNSARKSSSTVEIYVLMVDQÖRFDLATYINIMWS | 446 | |
| QY | 300 | GPVVALICMIYVITLLSPGTFGSAVFILFYFAMFASLTLYFPRKCAALDERÖKÖKN | 359 | |
| Db | 447 | APLOYLILALYLMWLGSVLAGVAVMLVPMVNAVMARKITTYOVAAHKKSKDNRTKLMN | 506 | |
| QY | 360 | EVLTYIKFIEKMYAKFASQSVÖKATREBERRILIEKAGYFOSITVGVAPIVVYASVYFS | 419 | |
| Db | 507 | ELINGIKVLYKLTAMELAPFDKVLAIROEELKAKKSATVLSAVGTFWTWCTPPLVALCTPA | 566 | |
| QY | 420 | VHMTGLFED--LTAÖAFVTVVYFNSKTRALAKTPRPSVKSLSASVAVDRFKSLFMEVH | 477 | |
| Db | 567 | VYVITDENNIIDAÖAFPSALFENLIRPPLNTLPMVISSIVASVSLKRLRFLSHEDLE | 626 | |
| QY | 478 | --MIKKEPASP---HIKTEKMNATLAMPÖSHSÖNSPRLPFRMKKDKRASKRÖKÖVNO | 532 | |
| Db | 627 | PSIERRRVKÖGGGINSITVRNATFTW----- | 653 | |
| QY | 533 | LÖRTEHQVLAÖKGHLLDSDERSPEEÖEKGHIHGLRLÖRTHSLIDLEIOÖCKLWG | 592 | |
| Db | 654 | -----ARSDPP-----TLNGITFSIPRGALVA | 675 | |
| QY | 593 | ICGSVGSÖKRTSILSAILDOMTLLBESSIASGPFYAVAOQWAILNMTIRNILLPGREYDE | 652 | |
| Db | 676 | VYÖGÖGCKRSLSLALMLÖMDKVEBHVAIKGSVAVAPQÖWAIÖNÖSLRNLITLFGÖÖLEP | 735 | |
| QY | 653 | RYNVYVNSCCLRPDALATPPSSDLTEIGERGAMVSGGÖRÖRISLARALYSDRSIYILDPL | 712 | |
| Db | 726 | YRYSYÖACALLPDLEIIPSGDRTEIGERGAVLSSGÖGÖRVSALARVYSNADYILEFDPL | 795 | |
| QY | 713 | SALDAAVGNHIFNSAI--RKHLKSTVLEFYVHÖLOLYLDCCDEVITPMKESCTIERGTHEEL | 770 | |
| Db | 796 | SAVDAAHVCKHIFENYIGRGMGLKNTRILLVTHSMSYLPOVDVIYVMSGKISEMSYÖEL | 855 | |
| QY | 771 | NMLNDGYATIF----- | 795 | |
| Db | 856 | LARDAAFEFLRTYASTÖQÖDÖAÖENÖYTVGSPGPEKÖKÖMÖNGMLVYDSAKÖLÖRÖLS | 915 | |
| QY | 796 | SKKETSGÖKSÖDKGPKTGSITKKEKAVKPEBGÖLVÖLEÖKGÖSGVPMVSÖVGYVIOAAG | 855 | |
| Db | 916 | SSSSYSGÖISIRHHN---STAEÖLÖKÖAEKKEFTWIKTMEADKÖTQGVÖKLSVYDMYDKAIGL | 972 | |
| QY | 856 | PLAFYVIMALEPLANGSRAFSTWMLSYMTKÖSGÖWTYVYTRGNETSVSDSMKÖNPMÖYXA | 915 | |
| Db | 973 | FISFELISIF--LEPCNHNVSALASMTWLSLMTDDEIVNGT---ÖEHYVRLSVIGALGISÖGI | 1028 | |
| QY | 916 | SIYALSNVMLILKAIÖRVVFEVYKGTILKASRSRLHDLFRILLSPKMFEDPTPTGRILNFE | 975 | |
| Db | 1029 | AVFÖGSMANVSI-----GGILASRCHVDLHLSILASPSMFERTPSGNMLVNF | 1076 | |
| QY | 976 | SKMDDEVURVLPÖFQAEÖMTÖQVILVFCVÖGMATGYCFPMFVLAVGÖPLVILFSLVLIHYSRYL | 1035 | |
| Db | 1077 | SKEDTÖVDSMPEVYIKTMÖGSLFENVIGACITVILATPTAAIITIPPLGILY---FVÖÖRY | 1133 | |

[illegible]

NAME: Steeg, Carol Miernicki
 REGISTRATION NUMBER: 39,539
 REFERENCE/DOCKET NUMBER: 01551
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (613) 545-2342
 TELEFAX: (613) 545-6853
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1531 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-460-907B-4

Query Match 28.5%; Score 2082; DB 2; Length 1531;
 Best Local Similarity 34.5%; Pred. No. 1,8e-195;
 Matches 488; Conservative 262; Mismatches 479; Indels 184; Gaps 21;

QY 100 PVDNAGLSCMTFSLASLAVAHKKKGLSMEDWVSLKSHSSDVNCRRLERLMQELNE 159
 DB 209 PESSASFLSRITFMWITGLT-VRGTRQPLEGSDMLNKEDTSEGVVPLVKNMKKECAK 267
 QY 160 V-----GPPAA-----SLRRVWIFCTRLIL 181
 DB 268 TRKQPVKVVYSSKDPQAKESSKYDANEVEALIVKSPQKEMNPRLFVLYKTFPGPYFLM 327
 QY 182 SIVCLMTQLAGFSGPAPMVKHLLEYTOATESNLOYSLVLGLLTETIVRSWLSALTWA 241
 DB 328 SEFFKAIDHLMFSGPQ-ILKLLIKFVNDTKAPDQGYFYVLLTFVTACLOTLVLYHQYFH 386
 QY 242 LNYRGTVRGAILLMARKILLKLNKEKS--IGELINICNDQORFEEAAVSSLAG 299
 DB 387 ICFVSGMKIKTAIVGAVRKALVITNSARKSTVEIYVLMASVDQRPMDLTVYINMWS 446
 QY 300 GPVVAIIIGMIVNVILPTGFGLSAVFIIFYPAMFASRLTAFFPKVCVAATDERVOKIN 359
 DB 447 APLOYTIALVLLMLNGLPSVLAVAVWVLMVAVNAVMAMKFTTYVAHMKSKDNKIKLMN 506
 QY 360 EVLYIYIKIKTAVYKASQSVOKIREBERILKEAGYFQSTITGVAPVIVVIVASVTFPS 419
 DB 507 EILNGIKVLKLYAMELAKEDKVLAIQOELKLYKKSAYLSAAGTFPTWCTPPLVALCTFA 566
 QY 420 VHMILGFD--LTAQAFIVVVFVNSMTFALKVTPSVKSSLSASAVADREKSLFMEVH 477
 DB 567 VYVITDENNIIDAOAFYSLALFLRLPPLNLLPVISSIVQASVSLKRLRFLSHHELE 626
 QY 478 --MIKNKPASP---HIKTEMKATLADSSSHSSSIONSPKLPFKMKDKRASGKKEVRO 532
 DB 627 PDSIFRRVVKDGGGNSITVRNATFTW----- 653
 QY 533 LQRTHQAVLAQKGLLLDSDERSPEEERKITHLGLRLQRTLSIDLEIQGKLVG 592
 DB 654 -----ARSDP-----TLNGITFIPGALVA 675
 QY 593 ICGVSGKTSLSAIILOMTLBESIAISGTFAYVAQOATLANLTLDNIIIFGEYDEE 652
 DB 676 VVGQVCGCKSSLSLALAEKMEKVEGHVAIKGSVAVPOAMQONSLRENILFICOLEEP 735
 QY 653 RYNSVYNSCCRLDALIPSDLTIEIGRGANLISGORORISLARATYSDRSIGIYLDPL 712
 DB 736 YRSVYQACALLPDLEILPSGDRTEIGEGVNLSSGOKORVSLAAVYNSNADIYLFDDPL 795
 QY 713 SALDAHVGNHIFNSAI--RKHLKSTVLEVTYHQQLVDCDEVIFMKBCGCTERTGTHEEL 770
 DB 796 SAYDAHVGHKHFENYIGKMLKNKTRILVTHSMGYLPQVDIIVMSGKISEMSGYGL 855
 QY 771 MNLNGDYATIF-----NNLL-----GETPPEVN 795
 DB 856 LARDGAFLEFLTYASTQEOADAENGVTGVSGPKKAKOMENGMLVYDSAKQLOLORLS 915
 QY 796 SKKETSQOKSODKPGTKSGIKKEKAVKPEEGOLVQLEKQGSVPMVSVYVYIOAGG 855

DB 916 SSSYSIGDISRRHN---STAELQKAEAKKEETWKLMEADKAQOTGVQKLSVYDYNKAIGL 972
 QY 856 PLALVIALALEPLANGSTAFSTWMLSYWIKQSGNTYTRRGNETSVSDSMKNPNMQYYA 915
 DB 973 FTSLSIF-LFCNHNVSALASVWLSLMTDPIYNGT---QERTVRLSVGALCISOCI 1028
 QY 916 SIYALSMVMLTKAIRGVVFEVKGTLIRASSRLHDELFRILRSPKFFDTPPTGRLNRF 975
 DB 1029 AVFGSMASVSI-----GGILASRCLHVDLHLSILRSPSPFERPSSGMLVNRF 1076
 QY 976 SKMDDEVVRLPEQAEMLTQNVILVFCVGMIAVGPMPVLAAGPLVILFSLHIVSRVL 1035
 DB 1077 SKELDTVDSMIEPVKIMKMGSSLFNVIGACIYVILLATPIAIIIPGLIY---FEVQREY 1133
 QY 1036 I---RELKRNITQSPFLSHITSSIOGATTAHANKQOELHRQOELLDDOARPEFLT 1092
 DB 1134 VASSQOLRLESVSHSPVYSHNETLLGVSIVRAFEQDERFTHQSDLVKVDENQAKVPSI 1193
 QY 1093 CAMRLAVRLDISALTITGMLIVLMHGQIPPAVAGIAISAVQLTGLFOFTVRLASE 1152
 DB 1194 VANRLAVRLKEVCNCIYLFALFAVISHSLASGLVGLSVSYQVTTYLMVLRMSSE 1253
 QY 1153 TEARTSVERRINHYIKTSLLEAPARIKKNKAPSPDPQEGEVTENAEKRYRENLPVLYLK 1212
 DB 1254 METNIVAVRELKEYSET-EKEAPMOIQETAPPSWPQYGRVFEFRYKCLYRREDLDFVLH 1312
 QY 1213 VSTFKPEKIGIVRTSGKSLGMALFRLVELSGGCKIDGVNISDGLADLSKLSI 1272
 DB 1313 INVTNGEKVGIVRTAGKSLTGLFRINESAGEGLIIDGINAKIGLHDLFFKTI 1372
 QY 1273 IPOEVLFGVYRSLNLPDNOYEDOIDALERTHMKKECIAOLPLKLESEVWENGDNFSV 1332
 DB 1373 IPDVPVFGSLRMLNDPSPQYSDSEEWTSLELAHKDQVSLPKLDHECEGGENSLV 1432
 QY 1333 GEROLLARALRLRCKILLIDEATAAMDTEFDLLIOETIRAFADCTMLTAHRLATVL 1392
 DB 1433 GORQLVCLARALLRKTKILLIDEATAAVDLETFDLDIGSTIRFQFDCVLTFAHRLNTM 1492
 QY 1393 GSDRTMIVLAQGVVEFDPSVLLSNDSSRFYAM 1425
 DB 1493 DYTRVIVLDKGEIOEYGAPSDLL-QOORGLFYSM 1524

RESULT 7
 US-08-463-179A-4
 ; Sequence 4, Application US/08463179A
 ; Patent No. 6001563
 ; GENERAL INFORMATION:
 ; APPLICANT: Cole, Susan P. C.
 ; APPLICANT: Dealey, Roger G.
 ; TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/463.179A
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/966,923
 ; FILING DATE: 27-OCT-1992
 ; APPLICATION NUMBER: 08/029,340
 ; FILING DATE: 8-MAR-1993

APPLICATION NUMBER: 07/966,923
 FILING DATE: 27-OCT-1992
 APPLICATION NUMBER: 08/029,340
 FILING DATE: 8-MAR-1993
 APPLICATION NUMBER: 08/141,893
 FILING DATE: 26-OCT-1993
 APPLICATION NUMBER: 08/407,207
 FILING DATE: 20-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Steeg, Carol Miernicki
 REGISTRATION NUMBER: 39,539
 REFERENCE/DOCKET NUMBER: 01547
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (613) 545-2342
 TELEFAX: (613) 545-6853
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1531 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-461-384B-4

Query Match 28.5%; Score 2082; DB 3; Length 1531;
 Best Local Similarity 34.5%; Pred. No. 1,8e-195;
 Matches 488; Conservative 262; Mismatches 479; Indels 184; Gaps 21;

QY 100 PVDNGLFSCMFFSMLSLARVHKKGELSMEDVMSLSKRESSDVNCRRLERLMQELNE 159
 DB 209 PSSASFLSRITFFWITGLI-VRGYRQPLEGSDMLNKEDISEQVVPVPLVNMKKCECK 267
 QY 160 V-----GPDA-----SURVWVIFCRRLIL 181
 DB 268 TRKOPVYVYSSKDPAPKESKSSVDANEVEALIVKSPOKENPISLFKVLKTFGPFILM 327
 QY 182 STVCLMITDLAGSGAPFVKKHLEETOATESNLQYSLVLGLLTETLVRSMLALMTVA 241
 DB 328 SFFFAIHDLMFSGFO-ILKLLIKFVNDTKAPDMOGYTYVLLVYACLOTLYLQYH 386
 QY 242 LNYRTGVLRLGAILTMAFKILKLNKERS--LGLINICSDGORMEAAVAGSLG 299
 DB 387 ICFVSGMRKTKTAVIGAVYKALVITNSAKSSVGEIVMLMSVDAQRFMDLATYIMMS 446
 QY 300 GPVVALIGITVYVILIGPFGISAVFIIFYPAMFASRLTAYFRKCYCAADDERQKNA 359
 DB 447 APLQVILALYLLMLNIGPSVLAGVAVMLVAVNAYMAKTKTYQVAHMKSKDNRIKLN 506
 QY 360 EYLITIKFIKMTAWKAFSQSVOKIREERRILEKAGYPOSTIVGAPLVIVIASVTFES 419
 DB 507 ELLNGIKVLKTLAMELAFKDKVALIRQELKVLKKSAYLSAVGTFWVCTPFLVALCTPA 566
 QY 420 VHMTLGFD-LTAAQAFVTVVFNMTFALKVTPFSVKSLSSEASVAVDFRSLFLMEVH 477
 DB 567 VYVTDENNIIIDAQRAVSLALFNIRPLNIIIPMWISSIVQASVSLKRLIRIFLSHELE 626
 QY 478 --MIKKRASP--HIKIDMKNTLAWDSSHSIONSPLKIPKMKDKRASRCKKRVKQ 532
 DB 627 PDSIERRPVKDGGTNSITVRNATFTW----- 653
 QY 533 LQRTHEQAVLAQKGLHLLDSDEPSPREEEGKHHLGHLRLQRTLHSLDLEQEKLYG 592
 DB 654 -----ARSDPP-----TLNGITFSIPBEALVA 675
 QY 593 ICGSVSGKTSILSALIGOMTLLEGISALISGTFAYVAQOAMILNATLRNIIIFGKRYDE 652
 DB 676 VVGQVCGGSSLSLALLAMDKVEGVAIKGSAYVPOQAMIONDSLRNIIIFGCOLERP 735
 QY 653 RYNSVNLSCCLRPDLALIPSSDLTEIGERGANLSGGORORISILARALYSRISYIILDDPL 712
 DB 736 YRSVYIQAQALLPDLIELIPSGRTEIGERGVNLSGGQKQVSLARAVYSNADIYLEDDEL 795
 QY 713 SALDAHVGHNIINSAL--RKHLKSKTVLVFTVHOLQYLVDCDEVIFMKEGCITERGHEEL 770

DB 796 SAVDAHVGHNIENVIIGPRGMLKNKTRILIVTHSMSTLPPQVDVYIYMSGKISMSGYQEL 855
 QY 771 MNLNGDYATIF-----NMLL-----GETPPEVIN 795
 DB 856 LARDGAFAEFLRTYASTBOQDAEENGVTGVSAGPKAEKQAMGMLVYDSAKQOLQROL 915
 QY 796 SKKETSQSKQODGPKTGSIKKEKAVPREGQVQLEKQOGSPWMSVYGYIQAAG 855
 DB 916 SSSSYSGDISRHHN--STAELOKAEAKKEETWKLEAKQOTGVQKLSVYDYMKAIDL 972
 QY 856 PLAFVIMALFMLNNGSTAFSTWMLSYWIKQSGNTTVYRGNETSVSDSKMNPMMQYVA 915
 DB 973 FTSFLSIF-LFMCNHYASALASVWLSMTDDIVNT--QETKVRLSVYCALGISQGI 1028
 QY 916 STYALSMAVMLILKAIKRGVYEVKGLTRASSRLHDELFRILNSPKMFPDTTPGRLNRF 975
 DB 1029 AVEGYSMAVSI-----GGILASRCLHVDLHLSLSPMSFFERTPSGNLVNRF 1076
 QY 976 SKDMDEVDVRLPQOAMFIONVILVFCVGMLAGVPEMFLVAVGPVLVLSVLIHIVRYL 1035
 DB 1077 SKELDVDSMIEPVTKMENGSLFNVIAGACIVILATPIAIIIPPLGLTY--FVQRPY 1133
 QY 1036 I---RELKRLDNITQSPFLSHITSSIOGLATIHAYNKGCEFLHRYOELDDNQAPFLFT 1092
 DB 1134 VASSRQKRLKLESVSRSPVYSHENETLIGSVYIRAFEDGERFTIHSOLKVDENQKAYPSI 1193
 QY 1093 CAMRLAVRLDLISALITTTGIMVIMHGQIPPAVAGLAISAVOGLFQFTVRLASE 1152
 DB 1194 VANRMLAVRLCEVGNCTIVFALFAVYSRHSIAGLVGSVSIQVTTYLNMVLVMSSE 1253
 QY 1153 TEARFTSVIRINHYITLSEAPARKNKAPSPDWQEBEYTFENAEVRYRELPLYLAK 1212
 DB 1254 METNIVAVEERLEKYESET-EKEAPWQIOETAPBSMWQVQVREVRNCLYRREDLVFLH 1312
 QY 1213 VSFPIPKKEKIGVGRGSGSGISGALFRLVELSGGCIKIDGVRISDGLADLRSLSI 1272
 DB 1313 INVTINGEKGVGIGRTGAGKSLTIGLFRINSAGEIILIDGINAKIGHDLRKRIT 1372
 QY 1273 IPOEVLFSGVYKSNIDPNOYTEDQIWDALERTHMEKCIQAOLPLKLESEVMEGDNFSY 1332
 DB 1373 IPDPPVIFGSLRMINDDPSQVSEDEEVMWLSLELAHKIDVSAIPDLIDECAGEGNLSV 1432
 QY 1333 GEROLLCTARALLRHCKIILIDBATAMDETDLIOETIRREAFACCTMLTAHRLTVL 1392
 DB 1433 GQROLVCLARALKRKTIYLVBATAVADLETDDLQSTIRIQTFOEDCTVITTAHRLTTM 1492
 QY 1393 GSDRIWLAQGVVEFDTPSVLLSNDSSRYAM 1425
 DB 1493 DITRIVLVLDKGEIQEYGAPSDLL-QQRGLEFYSM 1524

RESULT 9
 US-08-141-893-2
 ; Sequence 2, Application US/08141893
 ; Patent No. 5489519
 ; GENERAL INFORMATION:
 ; APPLICANT: Cole, Susan P.C.
 ; APPLICANT: Deoley, Roger G.
 ; TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEIN
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 STATE STREET, SUITE 510
 ; CITY: BOSTON
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,092B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: 01546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1531 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
OS-08-463-092B-2

```

| | | | | |
|---------------------------|--------|-------------------|-------------|--------------|
| Query Match | 28.48; | Score 2075; | DB 1; | Length 1531; |
| Best Local Similarity | 34.58; | Pred. No. 9e-195; | | |
| Matches 487; Conservative | 261; | Mismatches 481; | Indels 184; | Gaps 21 |

| | | | |
|-----|-----|--|-----|
| Qy | 100 | PVNAAGLESCMTEFWLSSLAIRVAHKKELSMEDVWSLSTKSHSSVYVNCRLERLMQOEINE | 153 |
| | | : | |
| Db | 209 | PSSASFLSTRLFTWITGLI - VRGYRPLGSDLSLNTKEDTSEQVVPVLVKKKKCCAK | 267 |
| | | : | |
| Qy | 160 | V - - - - - GPDA - - - - - SLRRVWIFCRLLIL | 181 |
| | | : | |
| Db | 268 | TRQRPVKVYSSKRPARESSKVDANBEVALLVKSPOKEWNSLTKVLKTFPEFLM | 327 |
| | | : | |
| Qy | 182 | SIYCLMTTOLAGFSGPAPMKHLEIYQATESNIQYLSLLYLGLLTLEIYRMSLALTMA | 241 |
| | | : | |
| Db | 328 | SFFKRAHDDLMFESGPO - IKLLIKFPNDKAPDMOGFYTVLLLEFVACQTLTLHOFH | 386 |
| | | : | |
| Qy | 242 | LNRYGVRLCRAITLMAFKILIKLNKIKES - LGEILINCSDNGRMEFAAAGSLIAG | 299 |
| | | : | |
| Db | 387 | ICFSGMKIKRAIVGAYYRKRLVLTNSARKSSYIGEYLNMSYDAQCFMDLATTIINNWS | 446 |
| | | : | |
| Qy | 300 | GPVAILGMIYVNIILGPTGFLSVAFTLEFPAMFASRLTAYFRRCVATDERVOYOMN | 359 |
| | | : | |
| Db | 447 | APLOVILALYLMINTLNGPSYLAGVAVVWLAMPVAAVAMKTKTYQVAMHMSKDRIRIKLM | 506 |
| | | : | |
| Qy | 360 | EVLVIYIKIKYAWVAKFSGVOKIREEBRIILKAGFOSITVGAPIVVLASVYTF | 419 |
| | | : | |
| Db | 507 | EILNGIKVLIKYNELLAKDKVLAIRDBELKVLKKSATLSAIVGFTWCTPFLVADLTFA | 566 |
| | | : | |
| Qy | 420 | VHMTLGED - LTAQAFTVYVNSMTFALKVTPBPYSKSSSEASVADVDRKSLFIMEVH | 477 |
| | | : | |
| Db | 567 | YVYVTDENNIIIDAOTAFYSLALFNILRFPILNIPWVISSIVQASVSLRILRIFLSHELE | 626 |
| | | : | |
| Qy | 478 | - - - MKNKRASP - - - HIKEMKNTLAWDSSHSSIONSPLKLPKKKKDKKRASRGKEVRO | 533 |
| | | : | |
| Db | 627 | PDSTIERVYKGGGCTNSTLYVNAFTW - - - - - | 655 |
| | | : | |
| 533 | | LQRTHEQVILAEQKGHLLSDERPPSPPEEBEGKHIHGLHRLQRTLSIDLEIOEGRLVG | 592 |

[illegible]

```

RESULT 11
US-08-462-109A-2
; Sequence 2, Application US/08462109A
; Patent No. 5682875
;
; GENERAL INFORMATION:
;
; APPLICANT: Cole, Susan P.C.
; APPLICANT: Deeley, Roger G.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: MULTIDRUG RESISTANT TUMOR CELLS

```


QY 1333 GEROLCTIARALLRHCKILILDEATAAMTETDILLOETIREAFADCTMLTIAHRLFTVL 1392
1333 GEROLCTIARALLRHCKILILDEATAAMTETDILLOETIREAFADCTMLTIAHRLFTVL 1392
Db 1433 GOROLVCLARALLRKTKILVLEDEATAVDELTDLQSIIRTOFEDCTVLTIAHRLNTIM 1492
1433 GOROLVCLARALLRKTKILVLEDEATAVDELTDLQSIIRTOFEDCTVLTIAHRLNTIM 1492
QY 1393 GSDRIWLAOGOVEFFTPSVLLSNDSSRYAM 1425
1393 GSDRIWLAOGOVEFFTPSVLLSNDSSRYAM 1425
Db 1493 DYTIVLVLDKEIOETIYGAPSDLL-QORGLFTYSM 1524
1493 DYTIVLVLDKEIOETIYGAPSDLL-QORGLFTYSM 1524
RESULT 13
US-08-463-179A-2
Sequence 2, Application US/08463179A
Patent No. 6001563
GENERAL INFORMATION:
APPLICANT: Cole, Susan P.C.
APPLICANT: Deoley, Roger G.
TITLE OF INVENTION: METHODS FOR IDENTIFYING CHEMOSENSITIZERS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
City: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,179A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Decont, Giulio A. Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: PQI-002CP8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1531 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-179A-2

Query Match 28.4%; Score 2075; DB 3; Length 1531;
Best Local Similarity 34.5%; Pred. No. 9e-195;
Matches 487; Conservative 261; Mismatches 481; Indels 184; Gaps 21;

QY 100 PYDAGLFCMCTEWSLSLARVAHKKGELSMEDVWSLSKSHSSDPVNCRRRLRLOEELNE 159
100 PYDAGLFCMCTEWSLSLARVAHKKGELSMEDVWSLSKSHSSDPVNCRRRLRLOEELNE 159
Db 209 PESSASFLSTRTTFMTWITGLI-VRGYRPLGSDLSLTKEDTSQVVPVLVKNKKKCEAK 267
209 PESSASFLSTRTTFMTWITGLI-VRGYRPLGSDLSLTKEDTSQVVPVLVKNKKKCEAK 267
QY 160 V-----GPDAA-----SLRRVWIFCRTRIL 181
160 V-----GPDAA-----SLRRVWIFCRTRIL 181
Db 268 TRKQPVKVVYSSKRPADPKRESSKYDANEVEALIVKSPQKEMNSLPKVLKTKTGPFILM 327
268 TRKQPVKVVYSSKRPADPKRESSKYDANEVEALIVKSPQKEMNSLPKVLKTKTGPFILM 327
QY 182 SYVCLMTITQLAGFSGPAPMVKHLETTQATESNLQYISLLVGLLLEIYVNSWSLALTWA 241
182 SYVCLMTITQLAGFSGPAPMVKHLETTQATESNLQYISLLVGLLLEIYVNSWSLALTWA 241

Db 328 SFFPKAHLDMFSGPD-LIKLLEKFEVNDTKAPDMQCYFTTVLLVYACIQTLVLYHOYF 386
328 SFFPKAHLDMFSGPD-LIKLLEKFEVNDTKAPDMQCYFTTVLLVYACIQTLVLYHOYF 386
QY 242 LMYRTVRLGAILITMAFKKILKLNKIKES--LGELINICSDGORMFEAAVGSLLAG 299
242 LMYRTVRLGAILITMAFKKILKLNKIKES--LGELINICSDGORMFEAAVGSLLAG 299
Db 387 ICFVSGMRITAVIGAVYRRALVITNSARKSSVYGEIVNLMSVDAQRFMDLATYINNIWS 446
387 ICFVSGMRITAVIGAVYRRALVITNSARKSSVYGEIVNLMSVDAQRFMDLATYINNIWS 446
QY 300 GPVVAILGMIYVNIILCPPTGFLGSAVFILFYPPAMFESRLTAIFRRKCVANAATEROVKN 359
300 GPVVAILGMIYVNIILCPPTGFLGSAVFILFYPPAMFESRLTAIFRRKCVANAATEROVKN 359
Db 447 APIQVILALYLLMLNLPGLSVLAGVAVVLMVNVAVAMVAKTKYQVAVHMSKDNRIKLM 506
447 APIQVILALYLLMLNLPGLSVLAGVAVVLMVNVAVAMVAKTKYQVAVHMSKDNRIKLM 506
QY 360 EVLTYIKFIMYAVWKAFSOSVOKIREERRILEKAGYFOSITGVAPIVVAVIASVYFS 419
360 EVLTYIKFIMYAVWKAFSOSVOKIREERRILEKAGYFOSITGVAPIVVAVIASVYFS 419
Db 507 EILNGIKVLYKIAMELAFKDKVLAIRQBELKVLKSKAYLAVAGFTVWCIPPLVALCTFA 566
507 EILNGIKVLYKIAMELAFKDKVLAIRQBELKVLKSKAYLAVAGFTVWCIPPLVALCTFA 566
QY 420 VHMTHGPD--LTTAAAFVTVVNSMTFALKVTPFSYKSLSEASVADRKSLFLEEVH 477
420 VHMTHGPD--LTTAAAFVTVVNSMTFALKVTPFSYKSLSEASVADRKSLFLEEVH 477
Db 567 VYVVIDENNLIDQATAVSLALFNILRPLNIPVVISIQAQSVSLKRLRIFLSHELE 626
567 VYVVIDENNLIDQATAVSLALFNILRPLNIPVVISIQAQSVSLKRLRIFLSHELE 626
QY 478 --MIKNKPASP--HIKIEKNATLADSSHSIQNSPKLTPKKKKDKRASRCKKEVYQ 532
478 --MIKNKPASP--HIKIEKNATLADSSHSIQNSPKLTPKKKKDKRASRCKKEVYQ 532
Db 627 PDSIERPPVMDGGGTNSITVYRNATFTW----- 653
627 PDSIERPPVMDGGGTNSITVYRNATFTW----- 653
QY 533 LQREHQAVALAEQGHLLDSDERSPEEBEKKIHHGLRLQRTLSIDLEIOEKLVG 592
533 LQREHQAVALAEQGHLLDSDERSPEEBEKKIHHGLRLQRTLSIDLEIOEKLVG 592
Db 654 -----ARSDPP-----TLNGITPSPIDEGALVA 675
654 -----ARSDPP-----TLNGITPSPIDEGALVA 675
QY 593 ICGSVSGKTSLSAILGOMTLLEGSLAISGTFAVVAQAMITLADNILEFKEDDEE 652
593 ICGSVSGKTSLSAILGOMTLLEGSLAISGTFAVVAQAMITLADNILEFKEDDEE 652
Db 676 VVGQVGGKTLSLALLAEMDKVGVHAIKGSVAVYQAQMIQDSLRNILEFCQLEEP 735
676 VVGQVGGKTLSLALLAEMDKVGVHAIKGSVAVYQAQMIQDSLRNILEFCQLEEP 735
QY 653 RYNSVLMSCIRPDLALPLPSDLTEIGERGANLGGORRISLARALYSDRSYIILDDL 712
653 RYNSVLMSCIRPDLALPLPSDLTEIGERGANLGGORRISLARALYSDRSYIILDDL 712
Db 736 YRSVYIACALPLDLEILPISGDRTEIGEKVNLGGOKORVSLARAYSNADYIIDLPL 795
736 YRSVYIACALPLDLEILPISGDRTEIGEKVNLGGOKORVSLARAYSNADYIIDLPL 795
QY 713 SALDAHGNHIFNSAI--RKHLSKTVLFWTHOLOVYLDDEVIFEMEGCTTERGTHEEL 770
713 SALDAHGNHIFNSAI--RKHLSKTVLFWTHOLOVYLDDEVIFEMEGCTTERGTHEEL 770
Db 796 SAVDAHGKHLFEENVIPKGMKLNKRLIVTHSMSTYLPQVDYIIVMSGKRISEMGIOEL 855
796 SAVDAHGKHLFEENVIPKGMKLNKRLIVTHSMSTYLPQVDYIIVMSGKRISEMGIOEL 855
QY 771 MNLNGDYATIF-----NNLLD-----GETPVEIN 795
771 MNLNGDYATIF-----NNLLD-----GETPVEIN 795
Db 856 LARDGAFAEFLRTYASTEQEDDAENGVTGVSFGPEAKQEMENMLVTSAGKQLQKOLS 915
856 LARDGAFAEFLRTYASTEQEDDAENGVTGVSFGPEAKQEMENMLVTSAGKQLQKOLS 915
QY 796 SKKETSQSKKSQDKPKRTGSIKKEKAVKPEEGOLVLEKGGQSVMSYGVYIQAAG 855
796 SKKETSQSKKSQDKPKRTGSIKKEKAVKPEEGOLVLEKGGQSVMSYGVYIQAAG 855
Db 916 SSSYSQSDISRHNH--STAELQKAEKAKKETWKLMDAKQOTQOVVLASYWDMYKAIIGL 972
916 SSSYSQSDISRHNH--STAELQKAEKAKKETWKLMDAKQOTQOVVLASYWDMYKAIIGL 972
QY 856 PLAFVIMALFMLNVGSPASTWMLSYWIKQSGNTVYRGNERSVSDSKNDPHMGUYA 915
856 PLAFVIMALFMLNVGSPASTWMLSYWIKQSGNTVYRGNERSVSDSKNDPHMGUYA 915
Db 973 FIFSLTIF-LFMCNHNVSALASNTWLSMTDDPIVNGT--QEHKVRLSYVAGALGISOGI 1028
973 FIFSLTIF-LFMCNHNVSALASNTWLSMTDDPIVNGT--QEHKVRLSYVAGALGISOGI 1028
QY 916 SIYALSMVMLILKAIGVVEFKGLTRASSRLHDELFRRLRSLPMKFEEDTPTGRILNR 975
916 SIYALSMVMLILKAIGVVEFKGLTRASSRLHDELFRRLRSLPMKFEEDTPTGRILNR 975
Db 1029 AVFGYSMAVSI-----GGLASRCLHVDLHLSIRSPMSFEFRPSCNLVNR 1076
1029 AVFGYSMAVSI-----GGLASRCLHVDLHLSIRSPMSFEFRPSCNLVNR 1076
QY 976 SKMDDEVDRILPQAEKFIQNVILVFCVGMIAGVFWFLVAVGLVILSVLHIVRVL 1035
976 SKMDDEVDRILPQAEKFIQNVILVFCVGMIAGVFWFLVAVGLVILSVLHIVRVL 1035
Db 1077 SKELDVTDSMIEVYIKKFMGSLFENVIGACIYIILATPIAIIIPPLGIY--FFVQRFY 1133
1077 SKELDVTDSMIEVYIKKFMGSLFENVIGACIYIILATPIAIIIPPLGIY--FFVQRFY 1133
QY 1036 I---RELKRLDNITQSPFLSIITSSIOGLATIHAYNKGQERLHRYGELLDONQAPFLFT 1092
1036 I---RELKRLDNITQSPFLSIITSSIOGLATIHAYNKGQERLHRYGELLDONQAPFLFT 1092
Db 1134 VASSRQRLKLESVRSRPSYSHFNETLIGSVYIRAFEEQERITHOSDLKVDNQAATYPSI 1193
1134 VASSRQRLKLESVRSRPSYSHFNETLIGSVYIRAFEEQERITHOSDLKVDNQAATYPSI 1193
QY 1093 CAMRWLAVRLDLISIALITTTGLMILMHQOIPPAVYAGLAIISYAVOLTLGFOFVRLASE 1152
1093 CAMRWLAVRLDLISIALITTTGLMILMHQOIPPAVYAGLAIISYAVOLTLGFOFVRLASE 1152
Db 1194 VANRWLAVRLLECVCNCIYLFALFAVISRHSLSAGVLGSLVSYQVYTYTINLVYRNSSE 1253
1194 VANRWLAVRLLECVCNCIYLFALFAVISRHSLSAGVLGSLVSYQVYTYTINLVYRNSSE 1253
QY 1153 TEARFTVERINHYIKTSLLEAPARINKAPSPDQEGEYTFENAMERRRENILPVKK 1212
1153 TEARFTVERINHYIKTSLLEAPARINKAPSPDQEGEYTFENAMERRRENILPVKK 1212
Db 1254 METNIVAVERLKEYSET-EKAPWQIOETRPPSSWPQVGRVEFRNYCLARYREDLDEFLVRH 1312
1254 METNIVAVERLKEYSET-EKAPWQIOETRPPSSWPQVGRVEFRNYCLARYREDLDEFLVRH 1312
QY 1213 VSPFIKPEKRTIGYGRGSGSSIGMALFRLVELSGGCIKIDGVRISDIGLADRSLKISI 1272
1213 VSPFIKPEKRTIGYGRGSGSSIGMALFRLVELSGGCIKIDGVRISDIGLADRSLKISI 1272

Db 1313 INVINGEKVIGVGTGAGKSSLTGLFRINESAGEIILIDGINIAKIGLHDLRFKITI 1372
QY 1273 IPOEVLGSGTVRSNLDPENQYTEDQIMDALERTMKECINADLPKLESEVENGDNVS 1332
Db 1373 IPDQVLTSGSRMLDPPSOYSDEEVTSLLELAHKDFVSALPKLHCEGEGENISV 1432
QY 1333 GEROLICARALIRKIKLILIDEAFAADTEFDLLOETIRAFADCTMLTAHRLHVL 1392
Db 1433 GORVLCARALIRKIKLILIDEAFAADTEFDLLOETIRAFADCTMLTAHRLHVL 1492
QY 1393 GSDRIMVLAQGOVEFDPPSVLLSNDSSRYAM 1425
Db 1493 DTRVIVLDKEIGIEYGA PSDLL-QORGLFYSM 1524

RESULT 14
US-08-461-384B-2
Sequence 2, Application US/08461384B
Patent No. 6025473

GENERAL INFORMATION:
APPLICANT: Cole, Susan P. C.
APPLICANT: Dealey, Roger G.
TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PARTEO RESEARCH & DEVELOPMENT INNOVATIONS
STREET: Queen's University at Kingston
CITY: Kingston
STATE: Ontario
COUNTRY: CANADA
ZIP: K7L 3N6

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,384B
FILING DATE: 05-JUN-95

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,923
FILING DATE: 27-OCT-1992
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Miernicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: 01547
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545-2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1531 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-384B-2

Query Match 28.4%; Score 2075; DB 3; Length 1531;
Best Local Similarity 34.5%; Pred. No. 9e-195;
Matches 487; Conservative 261; Mismatches 481; Indels 184; Gaps 21;

QY 100 PVDNAGLSCMTFSWLSLARIAHKKGELSMEDVWSLKHSSDYNCRLERLMDDELNE 159
Db 209 PSSASFLSRITFHWITGLI-VRGYRPLEGSDMLSNKEDTSEQVVPVLVNMKKECAK 267
QY 160 V-----GPDAA-----SLRRVVMIFCRRLT 181

Db 268 TRKQPVKVYVSSKDAQOPKSSKVDANEVEALIVKSPQKEMNPISLFVLYKTCGPYFLM 327
QY 182 STVCIMTIQIAGFSGPAFVWKHLEYTOATESNLOYSLIYGLLTELTVRSMSALATTA 241
Db 328 SFFFAFIHDLMMFSGPO-ILKLLIKFVNDTKAPDMQGYFYVLLVETVLAQTLVLYHQFYH 386
QY 242 LMYFVGRVRLGAILMAEFKILKLNKESK--LGLINIGSNDGOORMEAAVGSLLAG 299
Db 387 ICFVSGMRKTIIVICAVIKALVITNSARKSSTVEIYVLMVSDQREFDLATTYINMIS 446
QY 300 GPVVAILGNIYVNIILGPTGFLGSVFIIFYPAMMFASRLTAFFRKCAVATDEROVOKAN 359
Db 447 APLQVIALIYLMNLNMGPSVLAVGAVVMLVPMVNVMAKTKTYQVIAHKKSDNRIKLIN 506
QY 360 EYLYIKFTKMYAKVAKASQVOKIREERRLLEKAGFQSTITGVAPLVVVIASVYFS 419
Db 507 EILNIGIKVLKTLAMELAFKDKVLAIROBELKVLKKSAYLSAAGTFWVCTPPLVALCTFA 566
QY 420 VHMTLGFD--LTAQAFTVYVFNSTWTFALKYTPSVKSLSEASVAVDRFKSLFLMEEVH 477
Db 567 VYVTDENNILDAQTAFFVSLAFNLIRPPLNLPVYISIVASVSLKRLIFLSHEEL 626
QY 478 --MIRKPPASP---HIKIEKNATLWDSHSSSIONSPKLTFRMKKDKKRAKREKVRQ 532
Db 627 PSISERRPVKDDGGTNSITVRNATFTW----- 653
QY 533 IQRTEHQAVLABOKGHLILIDSDERPSPEEKGKHILHGLRLQRLHSIDLEIQEGLVG 592
Db 654 -----ARSDP-----TLNGITFSIPEGALVA 675
QY 593 ICGVSGSKTSLISALIGMTLLGSIASISGTFAYVAQOAMLTNRLNIFGKEYDE 652
Db 676 VVGQVCGCGSLSLALAMDMVKGVAIKGSVAYPQOAMONSLRENILFGCOLERP 735
QY 653 RYNSVLSNCCLPDLALIPSDLTEIGERGAMLSGQORISLARLYADRSIYIILDDPL 712
Db 736 YRSVYIACALLPDLIELPISGDRTELGEKGVNLSGQORVSLARVYNMAYILEDDPL 795
QY 713 SALDAHVGNHIFNSAI--RKHLKSKTVLFVTHOLOLYLVDCEVIEFKREGCITERGTHEEL 770
Db 796 SAVDAHVGHIFENYIGPGLMKKNTRLIVTHSMSTYLPQVDVIVMSGKISEMGSYQEL 855
QY 771 MNLNGDYATIF-----NNULL-----GETPVEIN 795
Db 856 IARDGAFAFLRTYASTEGQDAENGVTGVSQPKGAQOMENGMVLYTDSACKOLOROLS 915
QY 796 SKKETSGSQKSGODKPKGSIKKKAVKPEBGOLVQLEKGGQSVPMVYGVYIOAAG 855
Db 916 SSSSYSGDISRHHN---STAEIQKAEAKKEETWKLMEADKAQOTGOVKLSYVDYMKAIQL 972
QY 856 PLATVINALFNLNVGSAFTSTWMLSYWIKGSGNTVTRGNETSVDKMPHMYA 915
Db 973 FLSPLSLIF-LFMCNHVSALASNYWLSLMTDDPIVNGT---QEHYKRLVSLGALISQGI 1028
QY 916 STYALSMAVMLILKAIKRGVYFVKTRASSRIHDELFRILSLSPKAFPTTGTGRILNEF 975
Db 1029 AVFGYSMAVSI-----GGILASRCLHVDLHSILSPMSFFERTSPSGMLVNER 1076
QY 976 SKDMEVDVRLPFOAEMFIQNTVILVFCVGMITAGVPMFLVAVPVILFSLVLIHSRL 1035
Db 1077 SKELDTVDMSIPEVYIKMEGSLFPNVIACIVLILATPIAIIIPPLGLTY---FFVQRY 1133
QY 1036 I---RELKRLDNIQSPLSHITSSIOGLATIHAVNKGQEFLLHRYQELLDNOQAPFFLT 1092
Db 1134 VASSRQKLKLESVSSPYVSHNETLLVAGSVIRAFEGORRFTHQSDLKADENQAKAYPSI 1193
QY 1093 CAMRLAVRLDLISALITTTGMLYVLMHGOIPRAYAGIAISYAVQLRPLPFTYPLASE 1152
Db 1194 VANRWLAVRLCEVGNCIVLFAALFNVISRHSLASAGIVGSVSLOVTTVYLMWLVMSSSE 1253
QY 1153 TEARFTSVIRIHYYIKTSLSEAPARIKNAKAPSPDMVQOEVEYFEVNAEMRYRENPLVLAK 1212

Db 1254 METNIYAVERLKEYSST-EKEAPWQIOETNRPSSWPQVGRVEFPNRYCLRRREDIDFVLRH 1312

QY 1213 VSFITKREKIGIVRTSGSKSLGMALFRLVELISGCGIKIDGVRISDIGLADLRSLSI 1272

Db 1313 INTYNGEKGIVGIRGACKSSITLGLFRINBSABEIIIDINIKIGIHLDRFRITFI 1372

QY 1273 IPOEPVLFSGTNSNDPENQYTEDQIWDALERTHMECEIAQDPLKLESEVMEGDNFSV 1332

Db 1373 IPDPVLFSGSLRNNDLPFSQSDSEEVWSTLELAHLKDFVSALPDKLDHCEAGGEMLSV 1432

QY 1333 GEORLOLIAALRHCKIILIDETAMDTDLIOETIREAPADCTMLTIAHRLHTVL 1392

Db 1433 GQROLVCLARALLKRTIYLIDETAADVLETDLDLIQSTIRTFQEPDCTVLTIAHRLNTIM 1492

QY 1393 GSDRIMVIAOGOVEFDPDPSVLLSNDSSRFYAM 1425

Db 1493 DYTFRVIVLDKGEIOEGYCAPSDLL-QQGRLEFYSM 1524

RESULT 15

US-08-407-207A-2

Sequence 2, Application US/08407207A

Patent No. 6063621

GENERAL INFORMATION:

APPLICANT: Deeley, Roger G.

APPLICANT: Cole, Susan P.C.

TITLE OF INVENTION: ANTIBODIES TO A MULTIDRUG RESISTANCE PROTEIN

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: PARTRO RESEARCH & DEVELOPMENT INNOVATIONS

STREET: Queen's University at Kingston

CITY: Kingston

STATE: Ontario

COUNTRY: CANADA

ZIP: K7L 3N6

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/407,207A

FILING DATE: 20-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/966,923

FILING DATE: 27-OCT-1992

APPLICATION NUMBER: 08/029,340

FILING DATE: 8-MAR-1993

APPLICATION NUMBER: 08/141,893

FILING DATE: 26-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Steeg, Carol Miernicki

REGISTRATION NUMBER: 39,519

REFERENCE/DOCKET NUMBER: Q1512

TELECOMMUNICATION INFORMATION:

TELEPHONE: (613) 545-2342

TELEFAX: (613) 545-6853

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1531 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-407-207A-2

Query Match 28.4%; Score 2075; DB 3; Length 1531;

Best Local Similarity 34.5%; Pred. No. 9e-195;

Matches 487; Conservative 261; Mismatches 481; Indels 184; Gaps 21;

QY 100 PYDNAGLFCSTMTSLSTARVAHKKGLSMDYWSLSKSHSSDYVNCRRRLERLMOEELNE 159

Db 209 PESSASFLSRITFWITGLI-VRGYRQPLEGSDLSLNKEDTSEQVVPVLYKMKKECAK 267

QY 160 V-----GPDAA-----SLRRVYIFCRRLIL 181

Db 268 TRKQPVKVVYSSKRPADPKRESSKVDAVEEVALIVKSPQKEMNSLEKVIYKRTGPFYLM 327

QY 182 STVCLMTTOLAGFSGPAPMKHLEIYQATESNIQYSLLYLGLILEIYRSNLSALTMA 241

Db 328 SFEFKAIHDLMMFSGPO-IKLILIKFPNDKAPOMOGYFYVLLVFAACLOTLVLHOYFH 386

QY 242 LMYTGRALRGALITMAFKITLKNIKERS--LGEILINCSNDGQMFEEAAVSGSLAG 299

Db 387 ICFVSGRIKAVGAYGAYRRALVITNSARKSSITGEIYNLSVDAQFMDATYINNIWS 446

QY 300 GPVVALIGMIYVILIGPTGFLGSAVFLEFPYAMFASRLTAVRRRCVATDERVOKMN 359

Db 447 APLGVILLALYLMNLGPGSVLAGVAVVWLVPAVNAVMKTKTYQVAHMSKDRIKLM 506

QY 360 EVLTITKIKMYANVKAPOSQVOKIREERILEKAGYFOSITVGVAPVYVIVIASVTFS 419

Db 507 EILNGIVYKLKYLAMELFKDKVLAIRQELKVLKSAVLSAVGTFTWCTPEFLALCTFA 566

QY 420 VHMILGFD--LTAQAFTVTVFNSMTFALKVTPFSVKSLSSEASVAVDRFSLEFMEVH 477

Db 567 VYITIDENNIIDQAFAVYSLAFNLIRPLNLIIPWISSIYQASVSLKRLRIFLSHELE 626

QY 478 --MIKNRPASP--HIKEMKATLANDSSHSSIONSPKLTPKMKDKRASRGKREVRQ 532

Db 627 PDSIERPRVVKDGGTNSITVNAFTW----- 653

QY 533 LQTEHQAVLAEGKHLLDSDERPSPREEGRHIIHLRLQTLHSIDLEIOGKLYG 592

Db 654 -----ARSDPP-----TLNGITFSIPEGALVA 675

QY 593 ICGSVSGKSTLISAILGOMTLLEGSLAISTGFAYVAQAAILNATLNDLIFGKEDDE 652

Db 676 VVGQVGGCKSLSLALLAEMDKVEGHAIKGSVAVYVQAQMIQNDLSIRENTLFCQGLEP 735

QY 653 RYNSVLNSCCLRPDLAILPSSDLTEIGERGANISGGORQRISTARALYSDBRSIYLDPL 712

Db 736 YRSVIVQACALLPDELIPSGDREIGEKGVNLSCGQKQVRISLARAYSNADYILFDPL 795

QY 713 SALAHVGNHIFENAI--RKLKSKTVLFVTHQOYLVDDEVIFEMEGCTTERGTHEEL 770

Db 796 SAYDAHVGKHLFENVIGPKMLKNTKRLVTHSMSTYLPQVDVILVMSGKLSKSGSYOEL 855

QY 771 MNLGDAYATIP-----NNLL-----GETPPEIN 795

Db 856 LARGAFAPFLRTYASTEQEODAEENGVTGVSQPKKAKQKQENGMVLTDSAGKQLOLS 915

QY 796 SKKETSGQKKSQDKPKPTGSIKREKAVKPEGQVLEEKGGQSPWVSYYGYITQAGG 855

Db 916 SSSYSQGISRHNN--STAELOKAEAKKEETWKLMDADRKQOTQVLYSVWDMYKKAIGL 972

QY 856 PLATVIMALPMLVGSFAPSTWLSYVIKQSGNTYTRNENISYSDSMKDNHMOYXA 915

Db 973 FISFLSIF-LFCMHNVSALASNYLSMTDDPIVNGT--QEHKRVLSYVGAIGISQGI 1028

QY 916 SIYALSAVMILKAIKGVVFKTLRASSRLHDELFRIRISPMKFFDPTPTGRIINRF 975

Db 1029 AVFGYSMAVSI-----GGILASRCLHVDLHLSIRSPMSFEFRKPSGNLYNRF 1076

QY 976 SKMDDEVVRLPQAEEMFIQNVILVFCVGMIAGVFPWFVAVGPLVILSVLHIVSRVL 1035

Db 1077 SKELDVIDSMIPEVIKMKMGSLFNVIGACIVILLATPPLAIIIPPLIY--FFVQRFY 1133

QY 1036 I---RELKRLDNTQSPPLSIITSSIOGLATIHAYNKGQETLHYOELDNDQAPFLFT 1092

Db 1134 VASSROLKRLIESVSRSPYSHFNETLLGVSVIRAFEEQDERIHOSDKLVENQKAYVPSI 1193

QY 1093 CAMRWLAVRLDLISALITTTGMLIVLMHGOIPPAVYGLAITSVAVOLTFQFPVRLASE 1152

Db 1194 VANRWLAVRLECVCNCIYLPALPAVVISRHSLSAGLVGLSVSYSLQVTTYTNMLVRRSSSE 1253

THIS PAGE BLANK (USPTO)

| Result | No. | Score | Query | Match | length | DB | ID | Description |
|--------|-----|---------|-------|-------|--------|---------|----|---------------------|
| | 1 | 4811 | 65.8 | 946 | 1 | JC5667 | | multidrug resistan |
| | 2 | 23711.5 | 32.5 | 1437 | 1 | T20903 | | hypothetical prote |
| | 3 | 2308 | 31.6 | 1540 | 2 | T20904 | | hypothetical prote |
| | 4 | 2100.5 | 28.7 | 1541 | 1 | S71839 | | canalicular multidr |
| | 5 | 2075 | 28.4 | 1531 | 1 | DYHAR | | multidrug resistan |
| | 6 | 2068 | 28.3 | 1535 | 1 | S71841 | | multidrug resistan |
| | 7 | 2032 | 27.8 | 1495 | 1 | E86428 | | probable ABC trans |
| | 8 | 2023.5 | 27.7 | 1632 | 2 | D86428 | | hypothetical prote |
| | 9 | 1992 | 27.3 | 1633 | 2 | T01369 | | ABC transporter At |
| | 10 | 1983.5 | 27.1 | 1515 | 1 | S51863 | | cadmium resistance |
| | 11 | 1941 | 26.6 | 1537 | 1 | JE0336 | | canalicular multis |
| | 12 | 1928.5 | 26.4 | 1488 | 2 | F86328 | | probable ABC trans |
| | 13 | 1897 | 26.0 | 1573 | 2 | T21219 | | hypothetical prote |
| | 14 | 1882 | 25.8 | 1539 | 2 | T48059 | | ABC transporter-11 |
| | 15 | 1874.5 | 25.6 | 1478 | 2 | T348712 | | ABC transporter SP |
| | 16 | 1828 | 25.0 | 1355 | 2 | T00861 | | hypothetical prote |
| | 17 | 1828 | 25.0 | 1514 | 2 | T52080 | | multi resistance p |
| | 18 | 1824 | 25.0 | 1545 | 2 | T46645 | | sulfonylurea recep |
| | 19 | 1819.5 | 24.9 | 1516 | 2 | F84149 | | glutathione-conjug |
| | 20 | 1811 | 24.8 | 1515 | 2 | T52081 | | MRP-like ABC trans |
| | 21 | 1782.5 | 24.4 | 1545 | 2 | T424751 | | sulfonylurea recep |
| | 22 | 1781 | 24.4 | 1546 | 2 | T42728 | | sulfonylurea recep |
| | 23 | 1772.5 | 24.3 | 1450 | 2 | T47840 | | multi resistance p |
| | 24 | 1767 | 24.2 | 1144 | 2 | T27408 | | hypothetical prote |
| | 25 | 1749 | 23.9 | 1592 | 2 | S48933 | | probable transport |
| | 26 | 1741.5 | 23.8 | 1511 | 2 | T42711 | | sulfonylurea recep |
| | 27 | 1733.5 | 23.7 | 1559 | 1 | S64517 | | probable membrane |
| | 28 | 1727 | 23.6 | 1477 | 2 | S64616 | | YOR1 protein - yea |
| | 29 | 1723 | 23.6 | 1369 | 2 | T47796 | | ABC transporter-11 |

| | | | | | | |
|----|--------|------|------|---|---------|---------------------|
| 30 | 1713.5 | 23.4 | 1398 | 2 | T204343 | hypothetical prote |
| 31 | 1669.5 | 22.8 | 1582 | 2 | A56248 | sulfonylurea recep |
| 32 | 1665 | 22.8 | 1602 | 2 | T42216 | multidrug resistan |
| 33 | 1653 | 22.6 | 1421 | 2 | T34225 | hypothetical prote |
| 34 | 1649.5 | 22.6 | 1661 | 2 | S64800 | probable membrane |
| 35 | 1597.5 | 21.9 | 1037 | 2 | T50518 | ABC transporter-lik |
| 36 | 1521 | 20.8 | 1548 | 1 | DV4N5 | multidrug resistan |
| 37 | 1464 | 20.0 | 1446 | 2 | F84487 | probable ABC transp |
| 38 | 1454.5 | 19.9 | 1724 | 2 | T18343 | p-glycoprotein - S |
| 39 | 1451 | 19.9 | 1492 | 2 | A39322 | cystic fibrosis tr |
| 40 | 1450 | 19.8 | 1451 | 2 | T30821 | p-glycoprotein - S |
| 41 | 1417.5 | 19.4 | 1470 | 2 | T18344 | p-glycoprotein E - |
| 42 | 1408 | 19.3 | 1480 | 1 | DV4HCF | cystic fibrosis tr |
| 43 | 1393 | 19.1 | 1476 | 1 | A40303 | cystic fibrosis tr |
| 44 | 1386.5 | 19.0 | 1469 | 2 | T50210 | probable ABC trans |
| 45 | 1382 | 18.9 | 1453 | 2 | T26883 | hypothetical prote |

ALIGNMENTS

RESULT 1
JC5667
multidrug resistance protein, short type - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: JC5667
R: Suzuki, T.; Nishio, K.; Sasaki, H.; Kurokawa, H.; Saito-Ohara, F.; Ikeuchi, T.; Tan
Biochem. Biophys. Res. Commun. 238, 790-794, 1997
A:Title: cDNA cloning of a short type of multidrug resistance protein homologue, SMRP
A:Reference number: JC5667; MUID:97472289
A:Accession: JC5667
A:Molecule type: mRNA
A:Residues: 1-946 <SUZ>
A:Cross-references: DDBJ:AB005659; NID:q2554609; PTDN:BA422887.1; PID:q2554610
C:Genetics:
A:Gene: GDB:ABCC5; GRP5; SMRP; ABC33; MOAT-C
A:Cross-references: GDB:9954943
A:Map position: 3q25-3q27
C:Superfamily: human multidrug resistance protein, short type; ATP-binding cassette h
C:Keywords: ATP; nucleotide binding; P-loop
F:87-268/Domain: ATP-binding cassette homology <ABC1>
F:104-111/Region: Walker A motif
F:104-111/Region: nucleotide-binding motif A (P-loop)
F:214-219/Region: Walker B motif
F:719-912/Domain: ATP-binding cassette homology <ABC2>
F:736-743/Region: nucleotide-binding motif A (P-loop)
F:736-743/Region: Walker A motif
F:859-863/Region: Walker B motif

| | | | | |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match | 65.88; | Score 4811; | DB 1; | Length 946; |
| Best Local Similarity | 99.88; | Pred. No. 1.3e-315; | | |
| Matches 944; | Conservative 1; | Mismatches 1; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| QY | 492 | MKNNTLWADSSHSIIQNSPKITPKMKDKDRASRCKKEVRLQJTEHQAVLAEBQGHLL | 53 |
| Db | 1 | MKNNTLWADSSHSIIQNSPKITPKMKDKDRASRCKKEVRLQJTEHQAVLAEBQGHLL | 60 |
| QY | 552 | DSDERPSPSEEEEGKHILGHRLRORTLHSDLELQEGKLVGICGSVSGKTSLSIALIQ | 611 |
| Db | 61 | DSDERPSPSEEEEGKHILGHRLRORTLHSDLELQEGKLVGICGSVSGKTSLSIALIQ | 120 |
| QY | 612 | MTLLEGSIAIGTFAVYAQAQAMINATLROIILGKEYDERNSVYNSCCRLPDLAILP | 671 |
| Db | 121 | MTLLEGSIAIGTFAVYAQAQAMINATLROIILGKEYDERNSVYNSCCRLPDLAILP | 181 |
| QY | 672 | SSDLTEIGERGANISGGORORISLABALYSRSTYIIIDPLSALDAVHGNIIFMSAIRKH | 731 |
| Db | 181 | SSDLTEIGERKANISGGORORISLABALYSRSTYIIIDPLSALDAVHGNIIFMSAIRKH | 246 |
| QY | 732 | LKSKTVLEFVTHQLOLVDCDEVIIPKBECCITERTGTHELMNLNGDVATIFNNLLGETPP | 791 |

Db 241 LKSTVLFVTHQLOLVDCDEVIFMKECITERTGHEELMANGDIATFNNLLIGETPP 300
 Qy 792 VEINSKRETSQSKSODKPKRTGSIKKKAVKPEEGOLVLEEKQSGSPMSYGVYIQ 851
 Db 301 VEINSKRETSQSKSODKPKRTGSIKKKAVKPEEGOLVLEEKQSGSPMSYGVYIQ 360
 Qy 652 AAGGRLATLVLMALFMLNVGSAFSTWMLSVYIKQSGNTVTNRNENSVSDSMKDNPHM 911
 Db 361 AAGGRLATLVLMALFMLNVGSAFSTWMLSVYIKQSGNTVTNRNENSVSDSMKDNPHM 420
 Qy 912 QYASIVYLSMAVMILKAIGVYVVKGTLRASSRLHDELFRILRSPMKFFDTPTRI 971
 Db 421 QYASIVYLSMAVMILKAIGVYVVKGTLRASSRLHDELFRILRSPMKFFDTPTRI 480
 Qy 972 LNRFSKDMDEVDRPLPFOAEMFIQNVILVFCVGMIAGVFPWELVAVGPLVILFSLHIV 1031
 Db 481 LNRFSKDMDEVDRPLPFOAEMFIQNVILVFCVGMIAGVFPWELVAVGPLVILFSLHIV 540
 Qy 1032 SRVLLREKLRLDNTQSPFLSHITSSIOGLATIHAYNKGOEFLHRYQELLDNQAPEFLF 1091
 Db 541 SRVLLREKLRLDNTQSPFLSHITSSIOGLATIHAYNKGOEFLHRYQELLDNQAPEFLF 600
 Qy 1092 TCAMRWLAVRLDLISALITTTGMLIVLMHGOIPPAYAGLAISYVOLTGLFQFTVRLAS 1151
 Db 601 TCAMRWLAVRLDLISALITTTGMLIVLMHGOIPPAYAGLAISYVOLTGLFQFTVRLAS 660
 Qy 1152 ETEAFRTVERINHYIKTSLSEAPARIKKAPSPDQEGEVTEFENAEKRYRENTPLVYK 1211
 Db 661 ETEAFRTVERINHYIKTSLSEAPARIKKAPSPDQEGEVTEFENAEKRYRENTPLVYK 720
 Qy 1212 KVSFTIKRKERKIGYGRGSGKSSLGMLFLRVELSGGCIKIDGVRISDIGLADRSKLS 1271
 Db 721 KVSFTIKRKERKIGYGRGSGKSSLGMLFLRVELSGGCIKIDGVRISDIGLADRSKLS 780
 Qy 1272 IIPQEPVLFSGTVRSNLDPEFNOYTEDQIMDALERTHMEKCIQAOLPLKLESEVMENGDNFS 1331
 Db 781 IIPQEPVLFSGTVRSNLDPEFNOYTEDQIMDALERTHMEKCIQAOLPLKLESEVMENGDNFS 840
 Qy 1332 VGEROLLICARLRLHCKKILIDETAAMDPTDILLIOETREARADCMILTAIRLHTV 1391
 Db 841 VGEROLLICARLRLHCKKILIDETAAMDPTDILLIOETREARADCMILTAIRLHTV 900
 Qy 1392 LGSDDIMVLAOGVVEFDPSPVLLSNDSSRFYAFMAAENKAVK 1437
 Db 901 LGSDDIMVLAOGVVEFDPSPVLLSNDSSRFYAFMAAENKAVK 946
 RESULT 2
 20903
 hypothetical protein F14F4.3b - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T20903; T22217
 R:Wallis, J.
 submitted to the EMBL Data Library, January 1998
 A:Reference number: Z19344
 A:Accession: T20903
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1427 <WT>
 A:Cross-references: EMBL:AL021446; PIDN:CAB54225.1; GSPDB:GN00028; CESP:F14F4.3b
 A:Experimental source: clone F14F4
 R:Petcy, C.
 submitted to the EMBL Data Library, December 1996
 A:Reference number: Z19531
 A:Accession: T22217
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1427 <WT>
 A:Cross-references: EMBL:Z83227; PIDN:CAB54250.1; GSPDB:GN00028; CESP:F14F4.3b
 A:Experimental source: clone F45B8
 C:Genetics:
 A:Gene: CESP:F14F4.3b

A:Map position: X
 A:Introns: 26/2; 61/2; 116/2; 140/2; 218/2; 306/3; 356/1; 478/1; 534/3; 642/3; 669/3;
 C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homolo
 Query Match 32.5%; Score 2371.5; DB 2; Length 1427;
 Best Local Similarity 36.3%; Pred. No. 4,66-151;
 Matches 507; Conservative 289; Mismatches 497; Indels 103; Gaps 20;
 51 ETARAREGLSIDASMSHSLRLILDEHPKRYHHGLSALKPIRTTCKHOH---PYDNAGL 106
 26 ETSDRAKRYAGDTNKT-----IGRYSAVQNLPLTTERNNNGSGSRIDAGL 75
 107 PSCMTFMSLSIARAHKKGLSMEDVMSLSKHSSEVDNCRLELMOBELNEVPDAS 166
 76 FSEFTVSWFPLYQA-VRGKIDRNQWGCSPYDSCGLNMALELVWDE--KNAKAPS 133
 167 LRRVWIFCRFTLISIVCLMITOLAGFSGPAFWKHLELYQATESNLQ-----YSL 220
 134 LFKVYRFTSTLWFSCAVFFECILFGFICPTCFIRLLIAFENEPEDQSIYVYGLA 193
 221 LVGLLLEIYVRSMLATWALNRYGVRLGAILTMAFKILKIKERSLGLINIC 280
 194 LVAAISVEFAFVLSYGATMAVSYRTGIRVGAVLALLYKNVLNSKDLGKTESDVNIF 253
 281 SNDGORMEAAVAGSLIAGRPVALIGMIYVILIGPTGFGSAFILIYPRAMFASRLT 340
 254 ANDGRLDVAIFADPLVPLVGLVGGIGYLLMVLIGRMSLGLIFEVVDVYIOGELGSM 313
 341 AFRRRKVAATDERVOKMNEVLTYYIKFMAYVAKFASQYOKIREBRRLERKAGYQS 400
 314 VACRLALVYKTEKRISMMAEIIKYIRIVKMGWEQIFSAKIDQFKERKVOYRKSGYQS 373
 401 ITVGAPIVYVYASVYTSVHMTLGFDLTAQAFVYVFNMTALVTPPSVKSLSSEA 460
 374 LAIAGPVPVYVAALITFVGVALAGNDLASDAFSAIYVFYFMEIGIRIPIGYSYLSA 433
 461 SYAVDRFSLFMEEV--HMINKPASPIKITEMKNATLAMSSSHSSIONSFKLPKMKK 518
 434 VYAMRIQETYLLEOAYRYPVN---AEDVVLDCGATITTYO-----PK--- 474
 519 DKRASGRKKERYLOQRTENQAVLAEOKGHLLDSDERSPEEBGKHILHGLRLQRTL 578
 475 ---AAKAPVDETK--EPENEVIYVETPVF-----TC 501
 579 HSIDLEIDGKLVGICGVSQKTSLSAIILOQTLLEGSAIS--GTPAYVAQAWILNA 637
 502 SPDKLSIKRGEHIAVYGAVCGKSAIILKISGMFTTDALSDVRSQTVYVYOKAMIFNG 561
 638 TLRDNLIFGEYDEREYVNSLSCCLRPDLATLPSSDLTEIGERGANLSCGORISLAR 697
 562 TVQDNILFEDKKNSERYKAVANGCOLTEDTLLISGDKTEVEGERKATLSGOKAVALAR 621
 698 ALYSRSYIILDDPLSALDAVHGNHIFNSAIRKHLSKTVLFVTHQLOLVDCDEVIFMK 757
 622 AVEQTKNLYLFDIFASLDKRVANKIHEEIIQKLKKKALMVTNNMELHHFDRVLVE 681
 758 ECGITERTGHEELMANGDIATFNNLLIGETPPVPINSKRETSQSKSODKPKRTGSI 817
 682 GGNIVADGNHDLIERKNAYKTFVD-----ACEYQATSGATSPCGGAPQAPL 731
 818 KKE-----KAYKPEEGOLVLEE--KGQGVPMYSGYVYIOAAGPLAFVLMALFMLNV 870
 732 DAEIILNNSSEDLKGDADKILSDEEDMGNSTIMWRIYKOYIHAAAGMPITVCIGFIYAV 791
 871 GSTAFSTWMLSVYIKQ--SGNTVTNRGNE--TSVSDSMKDNPHMQYASIVYLSMAVMML 928
 792 VSNIFSTVLSMKLKKGHDETTITNGTELELMKTSIADSPTGYAAYVALVLTIS 851
 929 KAIRGVVYKGTLRASSRLHDELFRILRSPMKFFDTPTRIILNRSKDMDEVDRPLF 988
 852 GLFKACVFEVAVSLTAATRLHDMFOAVIHGATSPFDSPTGTGILNRFKSKDMEDIVKLPF 911

QY 989 CAEMTQVAVILVFCVCG---MIACGFFPFLVAVGCPVLVLESVLHVSVLIRELKRDLNI 1045
 Db 912 TAEVFLQWMT---CLGLVAVITSVFPEFLFAIPLFVAVFVESCFAIGIRNKKRSHI 968
 QY 1046 TQSPFLSHITSSIOGIATIHAYNKGQEFLLHRYOELLDNOAPFFLETCAMRLAVRDLI 1105
 Db 969 SRSPYLIDHVSASLEBITTIHFQOSNRFLVYLVKKHLDCNSCAIFEFOSAKMLAVMDLL 1028
 QY 1106 SIALITTYGLMIVLMHGOIPPAVAGIAISYAVOLGTLPFTVRLASEAFRTSVERINH 1165
 Db 1029 VYVMAIYVALLTVMLTGVSPADAGMALAFVQVSGIRQFVAVRTQTELEAKTISVERSY 1088
 QY 1166 YIKTISLEAPARINKAP-SPDWPEQCEVTFENAMEMRYRENPLVAKVSTFIKPKXIG 1224
 Db 1089 YADNIPPEGEMNTRGCLDISSWPANGQINFSEVWLRYRKSHPLANDITEIKKGEEYVG 1148
 QY 1225 IVGRGSGKSSLGMALEFVLELGGCICIKIDGVRISDGLADLRKSLTIIPQEPVFFSGTV 1284
 Db 1149 IIGRTGSKSSLANILFPLRYVTNGTIIYIDGVDLRTVGLVYLRGISAADPSSFSTV 1208
 QY 1285 RSNLDPEFQYTEDQIWDALERTHMKECIAQLPLKLESEVMGDNFVSGEROLCIARAL 1344
 Db 1209 RPNLDPSLEYSDSMITWEALERCHLKTIVQSLDKLEADVSHGNGNFSVGERQLPCLARAL 1268
 QY 1345 LRHCKILLIDRATAMDETDLLOETIREAFADCTMTIARLHTVGSQRIWLAOGQ 1404
 Db 1269 LMKSRIVLIDRATASVDAGTKLQEVIRKTFVADATVITIIAHLNDVNMRLMLKNGK 1328
 QY 1405 VVEFPTPSVLSSNDS 1420
 Db 1329 LINFTTPOEMERKDWMS 1344

 RESULT 3
 T20904
 hypothetical protein F14F4_3a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T20904; T22218
 R:Wallis, J.
 submitted to the EMBL Data Library, January 1998
 A:Reference number: Z19344
 A:Accession: T20904
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1400 <MT>
 A:Cross-references: EMBL:AL021446; PIDN:CAB54226.1; GSPDB:GN00028; CESP:F14F4_3a
 A:Experimental source: clone F14F4
 R:Pericy, C.
 submitted to the EMBL Data Library, December 1996
 A:Reference number: Z19531
 A:Accession: T22218
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1400 <MT>
 A:Cross-references: EMBL:T83227; PIDN:CAB54251.1; GSPDB:GN00028; CESP:F14F4_3a
 C:Genetics:
 A:Experimental source: clone F45B8
 A:Gene: CESP:F14F4_3a
 A:Map position: X
 A:Introns: 26/2; 61/2; 116/2; 140/2; 218/2; 306/3; 356/1; 478/1; 534/3; 642/3; 741/2; 776/2
 C:Superfamily: human multidrug resistance protein CMOAT2; ATP-binding cassette homology

 Query Match 31.6%; Score 2308; DB 2; Length 1400;
 Best Local Similarity 35.6%; Pred. No. 8.4e-147;
 Matches 497; Conservative 285; Mismatches 484; Indels 130; Gaps 21;

 QY 51 ETARAREGLSDASMSHSDRLIDEEHPKGYHHGSAKPIPTCKHOH---PVDNAGL 106
 Db 26 ETSDAKRYAGDTNKT-----IGRYSAVQNIPLPRTIRKNNKNGSKRIDDAGL 75

 QY 107 FSCMTFSWLSLARVAHKKGELSMEDVWSLSKHESSDVNCRRLERIMQDELINENVPDAAS 166

| | | | |
|----|------|---|------|
| Dd | 76 | FSEFYTSMVEPPLYQÅ-VRKGLDNRQWGCSEFSDSCGLMMARLEVMDEB-KKANAKSPS | 133 |
| Qy | 167 | LRRVWJFCRTRLLISIVCLMILITQLAGFSGPAPMVNHLLEYQÅVESNLO-----YSLL | 220 |
| Dd | 134 | LFKIYFIISIRLMFSCÅVVEFPLGIGFIPCFIRLLAFAPENBRDQSNIVYSTGIA | 193 |
| Qy | 221 | LVJGLLLEIVRSMSLALVWALNYFTGVRLRGAILTMAFRKILTKLNKIKESLUGELINIC | 280 |
| Dd | 194 | LVAISVVEFARVLVSATGAWAYSRTGIRVRGAVIALLTAKNLVNSKDLGCGSEVDINIE | 253 |
| Qy | 281 | SNDOGRPEAAVGSLLAGRPVAILIGMITYNIIIGPTFGSAPILFYPRAMMASLUT | 340 |
| Dd | 254 | ANDQRFEDATFAPLVLVGPRLVGVGIGYLLWIGRMSLSLIVFEPFDYIOFGLGFSM | 313 |
| Qy | 341 | AYFRKCAATDERVOXKMNVLTYFIKIKYUAVAKFASOVOKIREBERRLERKAGYOS | 400 |
| Dd | 314 | VACRNLAIIVTERKISMAELITIKIRLVKMNCHQJFSKIDQFKREKVQJRKSGYÅOS | 373 |
| Qy | 401 | ITVGVADIVVASVTEFSVHMTLGFEDLTAQAQFTVTVFNSMTFALKVTPFSVKSLEA | 460 |
| Dd | 374 | LAIACGPVPVPAALITFTFVGVVLAGMDLLASDFAISITVYFVMELGIRIMIPYGSRYLEA | 433 |
| Qy | 461 | SVAADREKSLFMEBIV--NHIKKRPASPHIKIEMKATLAMPSSHSSTIONSPLKPPKAK | 518 |
| Dd | 434 | VVARRRIOEYLLLEQVAPRYVTN--AEDVLDLQOATYTYQ-----PK---474 | |
| Qy | 519 | DKRASRGKKEKVRLOJTEHTEHOAVLEAOKGHLLDSDRSPREEBKNIHLGRLRQTL | 578 |
| Dd | 475 | ---AAKARVDETK--EPTENEVIVEPRV-----TC501 | |
| Qy | 579 | HSIDLEIOEGKLVGICGSVSGSTLSIALIGLQMTLLEGSIAIS-GTEPVYAQAQAMILNA | 637 |
| Dd | 502 | SFDKLSIKRGHEIAVIGAVCGSKSALIKAIISGHEMTTDDLASDSQUTYVVPQKAMFNG | 561 |
| Qy | 638 | TLRPNILTFKREYDERNVSNVSCCLRPDLALPSSDLIEIBRANISGGOORASLAR | 697 |
| Dd | 562 | TVQONLLIFGDKMNSERYKAVNCOQLTEDLITLVSQDTRVEGBRGATVSGGAKAVALAR | 621 |
| Qy | 698 | ALYSDRSYIILDDPLSLADAHVGNHIFNSAIRKHLKSKTYLFEVTHOLOLYVDCDEVIFMK | 757 |
| Dd | 622 | AVFQTKNLYFEDLIFASLDDKKLHLHF-----DVLVE654 | |
| Qy | 758 | EGCITERGCTHEELMNLNGDYATFNNLLGEPREVEINSKETSGSQKKSODKGRGTSGI | 817 |
| Dd | 655 | GNIVADGNHDIETEKNDAKTFEVD-----ACETYQATSGATSPCGGPAQAPPL | 704 |
| Qy | 818 | KKE-----KAVREBQOLVLEB-KGQGSVPMSVYGVYTOAAGGRLALVLMALFMLNV | 870 |
| Dd | 705 | DAELIRNSEDLLKQDARKLISDEDEMGNSITARIKYOIHAAAGMPITCVLGIPIVNV | 764 |
| Qy | 871 | GSTAFPMWMLSTYWKOG-SCNTVTYTGNE-TVSDSKMNQPHOYQVASYALASMAVMIL | 928 |
| Dd | 765 | VSNIFSTYWSRMLKXKHDETTTTITNGTEBLEKKTSLADSPYTGFAAYVALVALVYLTIS | 824 |
| Qy | 929 | KAIRGVVFKCTLRASSRLHDELFRILRSPMKFFDTPTGRILNFSKDMDEVVRLPF | 988 |
| Dd | 825 | GLFACVACEVKSILAAUVRDLRDMHQAVIHGATSFEDSTPGRILNFSKDMDEIVKLPE | 884 |
| Qy | 989 | QÅEMFQONVILVEFCVG---MIGVFPMPVLVANGPLVILFVSUHIYSRLVITELKRLDNI | 1045 |
| Dd | 885 | TAEFVLOHMIR---CLOFVLVITSVPEYFLFAPLFEVVFVVSQCFRAGIRNLRKSEHI | 941 |
| Qy | 1046 | TOSPLSITISIGGLATTIYANKGQEFRLRYOELDDNOAPFELTCSMRMLAVRLDI | 1105 |
| Dd | 942 | SRSPEDYDVHVSLESGITTTHTPOOSNRFEVLVKKHLDSCSALFPMQOSMRMLAVLIDL | 1001 |
| Qy | 1106 | SIALITTTGLMIVLMGQOIPRAYAGLISAYVOLTGLFOFVRLASSETBARTSVERINH | 1165 |
| Dd | 1002 | VVWMTAIVALLTVMLTGTSVRADAGMALFAVAGMSIIPQAVARTQTELEAKMTVEREVS | 1061 |
| Qy | 1166 | YIKTULSLEAPARIKNKAP-SPDMPQGBEVTFENAEKRYREBNILPVLYKVSFTIKREKIG | 1224 |

Db 1062 YADNIPEDGEMNTQOGDIIESWPAQINFSEVNLKRSKHPLANDITEIKGEKVG 1121
1225 IVGRFGSGKSSGLMALFRLVELSGGCIKIDGVRISDIGLADLRSLIIPQEPVLFSGTV 1284
1122 IIGRTGSGKSSGLMALFRLVELSGGCIKIDGVRISDIGLADLRSLIIPQEPVLFSGTV 1181
1285 RSNLDPENQVIEDDITWALFETTHKECIAOLPLKLESSEVMNGDNFSGEQQLCIAAL 1344
1182 RENDLPSTLESYSDSMWELECHTKTQVOSLDKLEADVSHSGNNFSGERQLEFLAROL 1241
1345 LRHCILILDEATPTEPDLIIQETIREAFADCTMLTIHRLHTVYAGSRIMVLAQO 1404
1242 LMKRIVYLDGATRSVADGPKLQIETVYKTFADATVYIIHRLDNRIMHKLNGK 1301
1405 VVEEDPSPVLLSNDSS 1420
1302 LINFTEPQEMFKDDMS 1317

RESULT 4
571839
canalicular multidrug resistance protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: S71839
R:Buechler, M.; Koenig, J.; Brom, M.; Kartenbeck, J.; Spring, H.; Horie, T.; Keppler, D.
J. Biol. Chem. 271, 15091-15098, 1996
A:Title: cDNA cloning of the hepatocyte canalicular isoform of the multidrug resistance
A:Reference number: S71839; MUID:96279006
A:Accession: S71839
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1541 <RUE>
A:Cross-references: EMBL:X63639; NID:91292881; PIDN:CAA65257.1; PID:91617207
C:Superfamily: human multidrug resistance protein CMOAT2; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
F:100-124/Domain: transmembrane #status predicted <TM01>
F:127-151/Domain: transmembrane #status predicted <TM02>
F:160-187/Domain: transmembrane #status predicted <TM03>
F:305-329/Domain: transmembrane #status predicted <TM04>
F:354-381/Domain: transmembrane #status predicted <TM05>
F:431-451/Domain: transmembrane #status predicted <TM06>
F:456-476/Domain: transmembrane #status predicted <TM07>
F:536-564/Domain: transmembrane #status predicted <TM08>
F:574-602/Domain: transmembrane #status predicted <TM09>
F:650-633/Domain: ATP-binding cassette homology <ABC1>
F:966-994/Region: nucleotide-binding motif A (P-loop)
F:1018-1046/Domain: transmembrane #status predicted <TM10>
F:1104-1132/Domain: transmembrane #status predicted <TM11>
F:1203-1228/Domain: transmembrane #status predicted <TM12>
F:1313-1506/Domain: ATP-binding cassette homology <ABC2>
F:1330-1337/Region: nucleotide-binding motif A (P-loop)
F:6,1007,1010,1011/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 28.7%; Score 2100.5; DB 1; Length 1541;
Best Local Similarity 34.4%; Pred. No. 9,1e-133;
Matches 497; Conservative 262; Mismatches 476; Indels 209; Gaps 26;

Db 98 OHPDNAGLSCMFEFMSLRLARVAHKGELSMEDVMSL-----SKHSS----- 142
191 QTSVTSVTSFSSITFSFSDRTVLKGYKH-PLTLEDVMDIDEGFTRSVTSKFEAMTKDL 249
143 ----DVCNRLLR-----LWQELINEVGPDAASLRVYVIF-- 174
250 QKARQARQRLQKQKRPKATLHGLNKKQSQDVLVELEKKKSEKTKRPKSWMLIKS 309
175 ----CRRLLIISVCMITOLAGSGPAFWKHLLEQTATSNLQYSLLVIGILLTEI 230
Db 310 LKFTFHVVLKSFILKLIHDLVFLNPQ-LIKLLIGVKKSSSYVMFGICAIIMFAVTL 368
231 VRMSLSLTLTMAINRTGVRLGALITLMAFKKILTKNI--KEKSLGELINICSDGQRMF 288

Db 369 IQSPCLQSYFOHCFVLGKCVFTYMSSTYKALTLISNLARKQYITGETVNLMSVDSQRLM 428
289 EAAVAGSLLAGPVVALIGMTYVNIILGPTGLGSAVFIILPYAPMMEFASRLTAVERKCV 348
429 DATNMOLVMSVQIITLSTIFFMRELQPSILAGVAVVLIIPNGVLAFTIRINIQVNM 488
349 AATDERQAKNEVLTLYKIKMAVAKFASQSVQKIREERRIIEKAGYFOSITGVAPI 408
489 KKKDKRLKIMNEIISGLIKLYKFAWEPFOBOGIRKEKELNLLRFQOISLFILOI 548
409 VVVLASVTFESVHTIGRD--LTAQAFTVTVFNMSFALKVTPESVKSSEASVAVDR 466
549 TPIIVSVTFESVTVLVDSANVLEKATSTILRNILFPLSMLEPWTSSILQASVADR 608
467 FKSLEFMEVMIKKNPASPRIK-IEMKNATLWDSHSSIQNSPKLTPKKKKDKRASRG 525
609 LERYIGDDDLTSAIRKSNPNDAKVKFEASFTMD----- 643
526 KKEVROLQRTQHOAVLAEOKGHLLDSDERPSPDEEKGIIHGLRLQRTLSIDLEI 585
644 -----PD-----LEATIDVNDI 657
586 QEGKIVGCGSVGSGKTSLSAIIIGOMTLEGSTAISGFAYVAQOAMILNATLNDITLF 645
658 KPGOLVAVGVTVSGKSSLSAMLGEMENVHGHITQGSTAVYVQOQVITQGTIKDNL 717
646 GKEYDEERYNVLNSCCRPDLALIPSSDLTEIGERGANLSGQORORISLARALYSRDI 705
718 GSEYNEKTYQVLAACALPLLELIPGDMATEIKGKINLSGQKORSLARAAYQADI 777
706 YILDDPSALDAHVGNHIFNSAIRKH--LKSRYLVFTVHQLOLYVDCDEVTKKECITE 763
778 YILDDPSAVAHVAKHIFNKVYVPGNGLACKTRIFVGHIFLPOYDEIVLVKGTILE 837
764 RGTHEELNUNGDAVATIPNNLLGETPVEI----- 794
838 KGYTRDLDDKKGVFARMKKTVMKHSRGEATVNNDSAEEDDDGLPTMEIPEDAASL 897
795 -----NSKET-SSGOKKQODKP-----KTSIKKEKAVPPEEGOLVLEKGGGS 840
898 AMRENSLRRLTSSRSSRSRSLKNSLTKIKVNVYLEKEKEVEGGKLIKKEFEVETGK 957
841 VPKSVYGYVIOAGCPALFVYLMALFPLNVGSTRFSTMVLSYMKQSGRNTVTRGETS 900
958 VKFSYLYKYLQAVGM-WSIFPIILFYGLNNAVFGISNLMISAMTSD--SDNI-----NGTN 1010
901 VDSNKKDNPQHGYASTYALSMAYVLLIKAIRGVVFGYKTL-----RASSRLHDELF 953
1011 NSSSHRD-----KRIQVFGALGLAOGICLLISTIMSYIACRNASKALHGOLLT 1058
954 RILRSPMKFEDTPTGRIINRFSKDMDEVYRLPQAEMLIQNVILVFCVG---MIAG 1009
1059 NILRAPMFFPTPTGRIVRNFSGDISFVDLLP----QTLRSMWMCFFGIAGTIVMCM 1114
1010 VEPWFVAVGVVLVFSVLAHVSRYVLRLEKRLDNITQSPLSHTSSIGLATIHAHYNK 1069
1115 ATPFAIILIPLSILYSVOFYVATSRQRLRDSYVKSPIYSHSEFVYGLPIITRAEH 1174
1070 GQEFLLHRYOELDDNOQAPFEFTCAMRLAVRLDISIALTTTGLMIVLHGGIIPAYA 1129
1175 QORLANNEKQIDNOKCVFSWITSNRRLAIRLELVGNLVFPGCALLIVYRKLTGLGDV 1234
1130 GLAISYAVOLGTFQVTVRLASSETARETSVERINHYIKTSLBAPARIKKNAPSPMPQ 1189
1235 GFVLSNMLNITQILNMLVRYMSSEATNIVAVERISEYI-IVENENAP-WYDVKRPADMPR 1292
1190 EGEVTFENAEKRYRENPLVYKKVSTTKPEKIGYIGRTSGSSGLMALFRLVELSGG 1249
Db 1293 HGEIOFNMYQYVRPELDVLKGTICNKSKEKGVVGRGAGKSLNLCFRILIESGQ 1352
1250 CIKIDGVRISDIGLADRSKSIIPQEPVLSGTVRSNLDPFNQYTEDQIDALERTHMK 1309

Db 1353 QIIDIGIVASIGLHDLERLITIPDPIIFSGSLRMNIDPENKYSDEBYRALFLAHR 1412
QY 1310 ECTIADLPKLESEVENDNFSVGEROLLCIARALLRCKIILIDEATAMDTEDLLIQ 1369
Db 1413 SFVSGIQLGILSEVEGDNDNISIGORQLCLGRVILRKSKILVLEDAFAVLEDTSLIQ 1472
QY 1370 ETRIRAFDCMTLTAHLHLTVLSGSDRIMVLAQGOVEDPFSVLLSDNSFEYAMF--A 1427
Db 1473 TTRKEFSQCCTVITTAHRLHITMDSKIMVLNDNGKIVEGSEPELLSNGS-FYLMARE 1531
QY 1428 AAEN 1431
Db 1532 GIEN 1535
RESULT 5
DYNHAR
multidrug resistance protein (cell line H69AR) - human
N:Alternate names: multidrug resistance-associated protein (MRP)
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence_revision 05-Dec-1998 #text_change 19-Jan-2001
C:Accession: A44231; A37495
R:Coile, S.P.C.; Bhargava, G.; Gerlach, J.H.; Mackie, J.E.; Grant, C.E.; Almqvist, K.C.;
Science 258, 1650-1654, 1992
A:Title: Overexpression of a transporter gene in a multidrug-resistant human lung cancer
A:Reference number: A44231; MUID:9308080
A:Accession: A44231
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-60 <CO2>
A:Cross-references: GB:L05628; NID:g1835658
A:Experimental source: small cell lung carcinoma cell line H69AR
A:Note: sequence extracted from NCBI backbone (NCBI:119851); this sequence has been cor
R:Coile, S.P.C.; Deeley, R.G.
Science 260, 879, 1993
A:Title: Multidrug resistance-associated protein: sequence correction.
A:Reference number: A37495; MUID:93262415
A:Accession: A37495
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-60 <CO2>
A:Cross-references: GB:L05628; NID:g1835658
A:Note: sequence extracted from NCBI backbone (NCBI:131929)
C:Genetics:
A:Gene: GDB:MRP
A:Cross-references: GDB:136335; OMIM:158343
A:Map position: 16p13.1-16p13.1
C:Superfamily: human multidrug resistance protein cMOAT; ATP-binding cassette homology
C:Keywords: antibiotic resistance; ATP; duplication; nucleotide binding; P-loop; transme
F:661-844/Domain: ATP-binding cassette homology <ABC1>
F:678-685/Region: nucleotide-binding motif A (P-loop)
F:788-792/Region: nucleotide-binding motif B
F:1310-1503/Domain: ATP-binding cassette homology <ABC2>
F:1317-1334/Region: nucleotide-binding motif A (P-loop)
F:1450-1454/Region: nucleotide-binding motif B

Query Match 28.4%; Score 2075; DB 1; Length 1531;
Best Local Similarity 34.5%; Pred. No. 4.7e-131;
Matches 487; Conservative 261; Mismatches 481; Indels 184; Gaps 21;
QY 100 PVDNAGLSCMTFMSLSLARAHHKKGELSMEDVMSLSKSHSSDVNCRLEKIMDEELNE 159
Db 203 PSSASFSLSRIFRWITGLI-VRGYRQPLEGSDLSLNKEDPSQVVPVLAIVNMKCECK 267
QY 160 V-----GPDAA-----SLRRVWIFCRRRLIL 181
Db 268 TKQKPVKVVYSSKDRAPQRESSKVDANEVEALIVKSPKEMNPISLFKVLKTFGTFLM 327
QY 182 STVCLMTQLAGSGPAFWKHLLEYTQATESNLOYSLDLVLGLLTETIVRSMLATWA 241
Db 328 SFFFKAHIDLMFSGPO-ILKLLIKFVNDTKAPDMQGVYFVLLFVTACTQLTLVHQYFH 386

QY 242 INTRGVRLRGALILMAPKILKLNKES--LGLINICNSDQORPEAAVGSLLAG 299
Db 387 ICEVSGMRKTAIVAGVARKALVTNSARKSSTVEIYMLKSVDAQREMDLVTYINMWS 446
QY 300 GVVVALIGIYVWVILIGPFGISAVFILFYPAAMFASRLTAFFRRKCAATDEROKAN 359
Db 447 APLQVILALYLMLNLGSPVLAVGVAVVLMVAVNVMAMKTTYQVAHMKSDNKLMLN 506
QY 360 EVLYIKFKIYKAWKASQSVOKIREERIRLEKAGFQSTIVGAVPIVVIASVTF 419
Db 507 EILNGIKVLKYAMELAFKDVLAIRQELKVLKRSAYLSAAGTFEWTCTPLVALCTFA 566
QY 420 VAMTIGFD--LTAQAFVYVYVFNMTFALKTPPSVKSLSASAAYORFKLFLMEEH 477
Db 567 YVYTIENNIIIDAQAFVSLALFNLRPLNLPVYISSIVQASVSLKRLIFLSHEEL 626
QY 478 --MINKPASP---HIKIKNATLAMDSSHSSSIONSPKLTFRMKKDKRASGKREKVO 532
Db 627 PSTIERRPYKDGCGNSTIVRNATFTW----- 653
QY 533 IQRTEHQAVALBQKHLLDSDRPSPEEESKNIHLGLRLQRLHSTIDLEIQEGKLVG 592
Db 654 -----ARDDP-----TLNGITFSIPGALVA 675
QY 593 ICGSGSGKTSILSALIGOMTLLEGSIAISGTFAYVAQOAMLTNRNILEFKREYDE 652
Db 676 VVGQVCGCGSLSLALMEDRVGHVALKGSVAAYVPOQAMQNSLRNILLFGQDLEP 735
QY 653 RYNSVLSGCLRPDLALIPSDLTEIGERAMISGQRORIASLARLYSDRSITYLDDLP 712
Db 736 YRSVIOACALLPDLFILPSCGRTIEGKGVNLGSGQKRVSLARVYSNADYILEDDLP 795
QY 713 SALDAHVGNHFNISAI--RKLHKSRTVLEVTTHQOLYVDCDEYIFKBCGTERGTHEEL 770
Db 796 SAYDAHVGNHFNENYIGRPMKLNKTRILVTHSMSTYLPQVDVILVMSGSKISEMSYQEL 855
QY 771 MNLNDYATIF-----NNLL-----GETPVEIN 795
Db 856 IARDGAFAEFLTYASTEOBDAENGVTGVSQPGKEAKOMEGMILVYDSAGQOLORUS 915
QY 796 SKKETSGSQKSDQKPKTKSIRKEKAVRPEGOLVLEKSGSVMSVGYVIOAGG 855
Db 916 SSSSYSGDISRHHN---STAELOKAAKKEEIKMLMEADKADQGVKLSVYDMYKAIIL 972
QY 856 PLATVYALFENLVNSTAFSTWMLSYWIKGSGNTVTRGNETSVSDSKNPNMQYTA 915
Db 973 FISTLSIF-LFMCNHYSAIASNYMLSLMTDDPIVNGT---QERTVRLSYGALGISQGI 1028
QY 916 STYALSMAYMLIKAIRGVVFKGTLRASSRLHDELFRILRSPMKFPPTPTGRILNRF 975
Db 1029 AVFGYSMAVSI-----GGILASRCLHVDLHLSILRSPMSFFERTPSGNLVNRF 1076
QY 976 SKDMDVDVRLPFOAEMFIQNTYLVEFCGMITAGYPMVLVAVPLVILFSLVHYRVL 1035
Db 1077 SKEDLTVDSMIEVYIKMFSGSLFENVIGACIVILLATPIAIIIIIPGLGLY---EFQORY 1133
QY 1036 I--BELKRLDNIQSPFSLTSSITIGIATIHAYNKGQELHRRQELLDNOAPEFLTF 1092
Db 1134 VASSRQLKLETSVSRSPVYSHFNETLLGVSIRAEEOBRFHQSOLKADENQKAYPSI 1193
QY 1093 CAMRQLAVRLDLISALTITTTGLMTVLMGQIPRAVAGIAISYAVOGLTGLPFTVRLASE 1152
Db 1194 VANRWLAFLVECGNCIVLPALPAVISHSISAGLIGLSVSTISQVTTYLMLVAMHSE 1253
QY 1153 TEARFSEVERINHYITSLLEAPARIKNKAPSPDMEQEEVEVFENAKMYRENPLVLK 1212
Db 1254 METNIVAAVERLKEYSRT-BEAPWQIOETRPSSWQVGRVERRNCTCLYRRDLDVFLH 1312
QY 1213 VSFITKPKKIGIVGRSGKSSLGALFLVLELGGCCKIDGVRSIDIGLADLRSKLSI 1272
Db 1313 INVYTINGGEKVGIVGRGKSSLTGLFRINESAGEELIYDINIAKIGLHDKRKITYI 1372
QY 1273 IPQEPVLFSGVYRSNMDPNQYTEQIMDALERTHKKECTIADLPKLESEVENDGNFVS 1332

Oy 1427 AAENKVAVK 1436
 | | : |
 Db 1535 AGIENVNSTK 1544

 RESULT 7
 E86428
 Probable ABC transporter [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E86428
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,
 Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hutzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maitl, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E86428
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1495 <STO>
 A:Cross-references: GB:AE005172; NID:g1055818; PIDN:AMG28288.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 A:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

| Query Match | 27.88; | Score 2032; | DB 2; | Length 1495; |
|---|-------------------|---------------------|-------------|--------------|
| Best Local Similarity | 34.66; | Pred. No. 3,6e-128; | | |
| Matches 486; | Conservative 240; | Mismatches 483; | Indels 196; | Gaps |
| QY 96 KHOHPVDNAGTGCSCMTGSMWLSLSTARVAHKGELSMEDWMSISKHESSDVNCRRRLERMAOE 155 | + | + | + | + |
| DB 224 EHICPERIASIFSNTYIGWITPLMQLOYKR-PIYERKYWOLDKMDQYETELIKRFRQKYTE 282 | + | + | + | + |
| QY 156 ELNEVGPDASLRRWVIF-----CRTLILSIVCIMITOLAGFSGPAEMVXHLLEYTO 209 | + | + | + | + |
| DB 283 E-----SRRPKPMLLRALNNSLSGRRFWLAGIFKIGDLSQFVGPPV-ILSHLLRSMQ 332 | + | + | + | + |
| QY 210 ATES---NLQYSLLLVIGLLTELTVRBSWLSLALYALYRGRVLRGAILTLMARKKIILK 266 | + | + | + | + |
| DB 333 EGDPAWGYVAFLIFEGVTLGVGICEQYFQNW---RQGFRLSTLVAIPRHSILRLT 388 | + | + | + | + |
| QY 267 NIKESKL--GELINICSDGORMPEAAVGSLLAGGVPVAIILGIVYVILLGPTFGLSA 324 | + | + | + | + |
| DB 389 HEARKNFRPSGVTWMTITDANALQOISQOLHGLMSAPFRITVSKILLYQOLGVAASLFGSL 448 | + | + | + | + |
| QY 325 VELLFPAMPASRLTAYFRKKCVAAATDEROVAKNEVLTYYIKFLMYAMKAFSOSVOKI 384 | + | + | + | + |
| DB 449 ILFLIPLQTLISKMRKRLREGLOMTDKRGITNELISSMDYVCKYAMKSPESRQGI 508 | + | + | + | + |
| QY 385 REEERRILKKA---GYQOSITVGAAPLYVYVYIASVYTSYMTKIGFDLTAAQATVTVYEN 441 | + | + | + | + |
| DB 509 RNEELSWERKQOLLSAVNSFLNSIPVY---TVSSGIVGLLGGDTLPRAAFSTSLFA 565 | + | + | + | + |
| QY 442 SMTEALKTPSPYSKLSSEASVAVDRFSLFMEVEVHIKKKPASPHI-KIEMKNATLAMD 500 | + | + | + | + |
| DB 566 VLRRPLNMLPRLLSQVYVNAVNSLQRIEELLSEPRILIAQNPRLPQPGPALISIKNGYRSM 625 | + | + | + | + |
| QY 501 SSHSIQNSPLLTPEKMKDKKRAKSGKKEKYVQLOQTEHQAVLAEQKCHLLDSDERSPE 560 | + | + | + | + |
| DB 626 S-----KTY----- 629 | + | + | + | + |
| QY 561 EEEGKHILHGLRLQORTLHSDLEIOGKLVGICGSGSKTSLISAILQOMLLE-GSI 619 | + | + | + | + |
| DB 630 -----KPTLSDINLEIPVGTLLAIVGGEKSTSLISMLDELSHAETTSV 675 | + | + | + | + |
| QY 620 AISGFAVAQAAMILNATLRDNLTFEKEYDEERYNSVLNSCCLRPLDIALIPSSDLTEIG 679 | + | + | + | + |

[illegible]

RESULT 8
D66428
hypothetical protein AAG28284.1 [Imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C.Accession: D66428
R.Thiologist: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.R.; Dewar,
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A.Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marzla,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D86428
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 11622 <STO>
 A:Cross-references: GB:AE005172; NID:g11055814; PIDN:AAG28284.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

Query Match 27.7%; Score 2023.5; DB 2; length 1622;
 Best Local Similarity 34.4%; Pred. No. 1.5e-127;
 Matches 478; Conservative 239; Mismatches 500; Indels 171; Gaps 21;

QY 74 EENHKGHHGLSLAKIRTTCKHQHVPYDANAGLFSCMTFSWLSLAVAKKGLMEDV 133
 DB 217 EELPGGE-----NICEPRH---ANLEDSIFFSWLNPMLTGLSKR-PLTEKDV 259
 QY 134 WSLKHESSDVNCRRLRLMOELENVEGPDAASLRVWVIFCPTRLLSIYCLMITOLAG 193
 DB 260 WHLDYMKETELMRSFQKSMDELEKRP-----W-----LIRALNLSLG 300
 QY 194 -----FSGAFVWKLLEYQTATES--NLQYSLLVGLLLEIVRSW 234
 DB 301 RFWVGFWKIGNDSCQFQV-LLNELLSQMLNEPAMIGYVAISIFVGVGLVCE-- 357
 QY 235 SLALTMLANTRGVRLGAILITMAFKILKLNKEKSL--GELINICSDNQMFEEAA 292
 DB 358 --AOYFQNMVWVGRILSALIAVERKSLRTNEGRRKFPQGTNTMTDAESLQIQ 415
 QY 293 VGSILAGPVVAIIIGPTGFLSAYVILFYPMAMFASRLTAVERRCVATD 352
 DB 416 SLHFMASPRRIYALVLIQGLGVAIIIGLFLVIMFPIQTVIISKTQKIKRGLQRTD 475
 QY 353 ERYOKNNEVLTYIKFIKYMVAKFASOSVOKIREERRILEKAGYFOSITGVAPIVVI 412
 DB 476 KRIGLMEVLLAAMPTKVCYAMENSFOGKQVTRDELSEWFKQQLLSAFNFIINSIPVL 535
 QY 413 ASVYTESVHMTIGDGLRAAQAFVYVFNSTFALKTPPSVKSLSASAVADVFKSLFL 472
 DB 536 VTVAVFQVFLSGDGLPARAFYSLSLFVSILFPLFPMIPITOMVANAVALNLEVLIS 595
 QY 473 MEVHMKNKPASP-HIKIEKNATLAWDSHSSIQNSPKLPKMKDKRASRGKKEVR 531
 DB 596 TEERVLLPNPPIEGGPAISIRNGYFSMD----- 625
 QY 532 QLORTEHQAVALBQKHILLSDERPSPEEKGKIHILGLRLQRTLSIDLEIOEGLV 591
 DB 626 -----KADRP-----TLSNINLIDPLSLG 645
 QY 592 GIGSGVSGKTSLSIALGLGM-TLLEGSIAISGTFVVAQOAMLTNRDNLIFGEYD 650
 DB 646 AAVGSTEGETKTSLSIAMIGELPARSDATVTLRGVSVAVPOVSWIFENATVRNIIIFGAPFD 705
 QY 651 EERNVNLNCCRLPDAIIPSSDLTEIGERGANLGGGORGORISIALVALYSDRSITLDD 710
 DB 706 QEKTERYIDYALQHDLELLPGDLEIGERGVNISGGQKRSMAKAVSNSPVCILDD 765
 QY 711 PLSLDAHVGNHIFNSAIRKHSKYVLFYTHOLOYVDDEVIFEMKEGCTTERGTHEEL 770
 DB 766 PLSLDAHVGOQYVEKIKRELIGOTTRVLTNQLHFLSYQDKILLVHEGYKEGTYEEL 825
 QY 771 MNINGDIATIFNNL-----LIGETPVEINSKKEPSSQOKSODKGPYTSIKKEKA 822
 DB 826 CHSGPLFQRLMENAGKVEDYSEENGAEVQTSYKPVENGANNMLQKDJETTKSKSGNS 885
 QY 823 VKRBEQGLVLEEGGSSVMSYGVYIOAAGPLAVLIVLALFMLVNGVSTAPSTWMLSY 882
 DB 886 Y-----LVKREKETGVSMKVLERYQNALGGAMVVMVLICVLTQVPRVSSSTWLSLSE 939

QY 883 WIKGSGNTVTRKNETSVSDSKMDNPHMOYASITVALSMAVMLILAKRGVFKGLR 942
 DB 940 WTDGSG---TPKTHG-----PLEYNNIVYALLSPGVSVLLINSYWLIMSLSLX 982
 QY 943 ASSRLHDELFRILRISPMKPFEDTPTGRILNRSKMDDEVDVRLPFOAKEMIOVILVVF 1002
 DB 983 AAKKMDMLGSLIRAPVFFQTPPLGRITRKFAKMDGDDIRYAVVYNNMFGSIAQLLS 1042
 QY 1003 CVGMIAGVFWFLVAVGPLVILFVSLHIVSRVLRLEKRLDNTIQSPFLSHITTSIOGLA 1062
 DB 1043 TVILIGIVSTLSMAIMPLVVFVGVATLYYQNTSREIKRMDSTRSPPVYQFGALNGLS 1102
 QY 1063 TIHAVNKGQEFLLHRYQELLDNQAPFLETCAMRMLAVRDLISALITTTGLMTIVMHG 1122
 DB 1103 SIRAYKAYDRMAEINGRSMNINIFTLVNMAMNMLGIRLEVLGLMVLASLAVMQNG 1162
 QY 1123 QI--PRAVA---GLAISAVOGLTGFQTVYVLAETEARFTSVSRINHYIKTSLSEAPAR 1177
 DB 1163 KAANOAASTMGILLSTALSITSSLRVLRKLSLAENSLNSVERGVNTEIPS-EAPLV 1221
 QY 1178 IKNRPSDPMQEGEVTEFENAEKRYRENLPVLKRVSTTIKPKKEKIGVGRTGSGKSLG 1237
 DB 1222 IENRPPGMPSSGSISFEDVLYLRPELPVYLGVSFLSPMDKVGIVGRTAGKSSLL 1281
 QY 1238 MALRLVELSGGCIKIDGVRISDGLADLNSKLSIIPQEPVLSGTVRSNIDPPNQYTED 1297
 DB 1282 NALFRIVLELEKGRILIDECDIGRGLMDLRKRVLGIIIPQAPVFGTVRFNIDPSEHND 1341
 QY 1298 QIWDALERTHMKCEIAQPLKLESEVMENGDFSVGEORGLCIRALLRHCKITILDEAT 1357
 DB 1342 DWESLERAHKDKTIRRRPGLDAEVTBAGENFSVGQRLSLARALLRKIKIIVDEAT 1401
 QY 1358 AAMDTELDLIQETIREAFADCTMLTIAHRLTVLGSDBRMVLAQOGVEEDFSPVLLSN 1417
 DB 1402 AAVDVRDVLIIQKTIREEKSCMLIIAHRLNTIIDCKVVLDSGKVGQERSPENILSN 1461
 QY 1418 DSRFPYAM 1425
 DB 1462 GESSFSKVM 1469

RESULT 9
 T01369
 ABC transporter AtMRP2 [imported] - Arabidopsis thaliana
 N:Alternate names: multidrug-resistance protein homolog T29F13.13
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
 C:Accession: T01369; D84759
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
 submitted to the EMBL Data Library, May 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.
 A:Reference number: 214179
 A:Accession: T01369
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1623 <ROU>
 A:Cross-references: EMBL:AC003096; NID:g3132469; PID:g3132479
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A>Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487
 A:Accession: D84759
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1623 <STO>
 A:Cross-references: GB:AE002093; NID:g3132479; PIDN:ANCL6268.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g34660; T29F13.13
 A:Map position: 2

A: Introns: 110/3; 170/2; 183/3; 236/1; 274/2; 301/2; 309/3; 328/3; 350/3; 375/3; 411/3;
C: Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology
F: 125-1452/Domain ATP-binding cassette homology <ABC>

| Query Match | 27.3% | Score 1992: | DB 2: | Length 1623: |
|-----------------------|-------------------|--|-------------|--------------|
| Best Local Similarity | 34.7% | Pred. No. 2e-125: | | |
| Matches 478: | Conservative 231: | Mismatches 487: | Indels 182: | Gaps 27: |
| QY | 100 | PVDNAGLPTSCMTFSWLSLSARAHKKGELSMEDVSLSKHESSDVNCRLERLMOEOLNE | 159 | |
| Db | 227 | PEKHNINFDKLFPSMNMNLMTGSKR-PLTEEDVWYLDTWOTELLFTFSFGSKMKELOK | 285 | |
| QY | 160 | VGPDAASLRVYVIFCRRLILSIYCLMITOLAG-----ESGPAF--M | 200 | |
| Db | 286 | POP-----W-----LRLALNNISGGFRFWMGFEWKGIGNDCSQFYGPLLNLDT | 326 | |
| QY | 201 | VKHLLLEYQATESNLQYSLILVGLLLEIYVSWSLATLWALNTRGVRLRCALITLMARK | 260 | |
| Db | 327 | LKSMEDAPAMMGYI-YAFSIFGVYFVGLCE-----AQYFQVNRKGYRLSRBALIAYNR | 381 | |
| QY | 261 | KILKIKNIKESL--GELINICSDNGQRMFEAAVAGSLLAGPVVAIGIYVNIILGPT | 318 | |
| Db | 382 | KSRLRTNGRRKRFQGTITNMTLTDAESLQICQSLHTWMSAPFRILIALILLYOOLGVA | 441 | |
| QY | 319 | GFLGSANVILLEYPAAMKFSRLATYRRRCVATDERVOKMNVLYIKIRIKYANWAKS | 378 | |
| Db | 442 | SLIGALLLVLMFPLOTYIIISKMKLTKGLOTRDRIKIGMENVLAAMDVVKCYAEMNSFO | 501 | |
| QY | 379 | QSVOKIREERIRLEKAGYFOSITGVAPYIVYVIAASVYFVSVMHMLGFDPLTAOAQVTVT | 438 | |
| Db | 502 | SKVOYVRDELDSWFRKSSQLGLANMFILNISIVLYTISFYGFYTLGGDLTPARAFYTS | 561 | |
| QY | 439 | VFNSTFALKVTPESVKSLSSEASVAVDRFKSLTFLEEMHMIKKNPASP-HIKIEMKNATL | 497 | |
| Db | 562 | LEFAVRFPFLFMPJNITTOVNVANVYSLKRLIEVLAITEERILLNPNPIEPGPAISIRNGYF | 621 | |
| QY | 498 | AMDSHSSIONSPLTKPKMKDKRASRCKKEKVVROLQRHNAVLAEQGHLLDSDEP | 557 | |
| Db | 622 | SMDS-----KG-----DRP | 630 | |
| QY | 558 | SPEEEGKHILHGLRLORTLSIDLEIOEGKLIVICGSVSGSKTSLISAIIQOMTLLEG | 617 | |
| Db | 631 | -----TLSNILDVPLGSLAVVSGTGKSKTSLISAIIELPATSD | 671 | |
| QY | 618 | SI-AISGFFAYAAQOAMTLNATLRNIIIFPKGTYDEBRINSVINSCLLRPDIAILPSSDLT | 676 | |
| Db | 672 | AIWTLRGSVAAVPPQYSWIFNATVNRNIIIFSGPFRDEKYEIRADIVTSLIKHDELTEPGCDLT | 731 | |
| QY | 677 | EIGEENALMSGGQRORISLARLYSDRSYIILDDPLSALDAHVGNHIEFSAIRKHLKSKT | 736 | |
| Db | 732 | EIGEENAVISGGQKORVEMARAVYNSDVTYFPDDPSIALDAHVGOQVFEKCIKRELQKQT | 791 | |
| QY | 737 | VLFVYHQQLVYDCEVIFPMKEGCTITERGTTHEELMLNLDYATIFINNLIG----- | 787 | |
| Db | 792 | RVLVYLNQHLFSLQVDRILYVHGGYTKEGSTYELSS-NG---PLQRLMENGKVAEYSE | 847 | |
| QY | 788 | ETPVEYINSKKE-----TSGSQKK-SODGPKTKGISIKKEAKVAPREBQOLVLEEKGGQ | 839 | |
| Db | 848 | ENGGEAEADQTEQPPANGNTNGIOMDGSDDKSKSGKNKGKGSV-----LIKQERETG | 901 | |
| QY | 840 | SVPMGVGVYVIOAAGCPPLAFIVMLMLFNMGSTAFSTWMLSWYIKQSGNITVIRGNT | 899 | |
| Db | 902 | VYSNRYVLRKIODALGAGANVWMLLCYVLTVEYFRYTSWLSSEWMDAG---TPKSHG--- | 955 | |
| QY | 900 | SVSDSMKDNPMHQYASIVYALSMAVMVILKAIKRGVVFVGTPLRASSRLDELDEFRRLLRSP | 959 | |
| Db | 956 | -----PLFNYLIYALLTSFGQVHLYTLNYSWMLMSLSYAAKRLHDMNLSILRAP | 1004 | |
| QY | 960 | MKFPPTTTPGRLNFSKDMDEVDYRLRPQAMFQNIQVILVFFCYGMATGAVRPWFLVAVG | 1019 | |
| Db | 1005 | MSFETHNPLGRINLEFADLDGIDRTFVAVFMMEGVQVQSLLSTAVLLIGIVSTLSIMAT | 1064 | |

| | | | |
|----|------|---|------|
| QY | 1020 | PVLIFSVLHVSVLRLREKLDINTDOSPESHITSSIQGLATIHVKNQGEFLHRQ | 1079 |
| Db | 1065 | PLVLFGVALYQNVYARERKKRMDSSRSRVPYAQFGEALINGSTIRAK-----AYDR | 1117 |
| QY | 1080 | LLD-----DNQAFPLEFTC-AMRWIAVRDLISILITTTGMLIVLMHQI--P'PAAV- | 1129 |
| Db | 1118 | MADINGSRMDNNIRFTLVNMGANRWIGIRLEFTJGGLMILWLSFVAMQGRARENQOAR | 1177 |
| QY | 1130 | --GLAISYAVOLTGLFOFTVRLASSETEARFTSYERINHYIKTLSLEAPARIKNKAPSPDW | 1187 |
| Db | 1178 | TGMLLSTALNTITSLTGLVRLASISLANSINAVERGANYIE-IPPEAPVIEINNRPPG | 1236 |
| QY | 1188 | POGEVYTFENAEYRRENDLVLYKAYSTFKKPEKIGIYGRGSGSSGMLAFLFYLS | 1244 |
| Db | 1237 | PSSGSKFEDVLYRQOLPPVYLGVSFFLHPDKVGIVGRGAGSSLLNLFRIAYE | 1296 |
| QY | 1248 | GGCJIKIDGVRISDGLADRSKSLIIPQEVLTSGSVRSNLDPFNQYTDQJWDLAERTH | 1307 |
| Db | 1297 | KGRILIDDCVGKRGMLDKRYGLIIPQSVLTSGSVRRNLDPFGHNADJWELERAH | 1356 |
| QY | 1308 | MKECTAOLPLKLESEVMGNDNFSVGEROLLCTARALLRCKLILIDEATAADTETDL | 1367 |
| Db | 1357 | LKDTIRRPDGLIDAEVSEAGENSVCGRQLSTLRALLRSRKILVDEATAAVDRTAL | 1416 |
| QY | 1368 | IOETIRREAFACCTMLTIAHRLHVYLSDRIMVLAQOQVYEFDPSTVLLSNSDSRYAM | 1425 |
| Db | 1417 | IOKTRREFKCTMLIIAHRLMIIIDCKLIVLDSRGVOEFSSPENLLSNBSSFSKM | 1474 |

RESULT 10

S51863

cadmium resistance protein YCF1 - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YD9302.11c; protein YDR135C

C:Species: *Saccharomyces cerevisiae*

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: S51863; A55352; S50233

R:Oliver, K.; Harris, D.

submitted to the EMBL Data Library, February 1995

A:Reference number: S51853

A:Accession: S51863

A:Molecule type: DNA

A:Residues: 1-1515 <OLI>

A:Cross-references: EMBL:Z44179; NID:9665657; PIDN:CA488217.1; PID:9665668; GSPDB:GN0

R:Schzyzpa, W.S.; Wemmie, J.A.; Moye-Rowley, W.S.; Thiele, D.J.

J. Biol. Chem. 269, 22853-22857, 1994

A:Title: A yeast metal resistance protein similar to human cystic fibrosis transmembr

A:Reference number: A55352; MUID:94357936

A:Accession: A55352

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-679, 'R', 681-1515 <SZC>

A:Cross-references: GB:I35237; NID:g556464; PIDN:AAA50353.1; PID:g556465

C:Genetics:

A:Gene: SGD:YCF1; MIPS:YDR135C

A:Cross-references: SGD:S0002342; MIPS:YDR135C

A:Map position: 4R

C:Function:

A:Description: required for cadmium resistance

C:Superfamily: human multidrug resistance protein CMOAT2; ATP-binding cassette homolo

C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein; yeast vacuole

F:287-308/Domain: transmembrane #status predicted <TM1>

F:345-366/Domain: transmembrane #status predicted <TM2>

F:421-442/Domain: transmembrane #status predicted <TM3>

F:446-467/Domain: transmembrane #status predicted <TM4>

F:534-555/Domain: transmembrane #status predicted <TM5>

F:558-580/Domain: transmembrane #status predicted <TM6>

F:646-829/Domain: ATP-binding cassette homology <ACCI>

F:663-670/Region: nucleotide-binding motif A (P-loop)

F:951-972/Domain: transmembrane #status predicted <TM7>

F:1068-1088/Domain: transmembrane #status predicted <TM8>

F:11092-1113/Domain: transmembrane #status predicted <TM9>

F:1119-1200/Domain: transmembrane #status predicted <TM10>

F:1119-1200/Domain: transmembrane #status predicted <TM11>

F:1208-1229/Domain: transmembrane #status predicted <TM12>
 F:1289-1463/Domain: ATP-binding cassette homology <ABC2>
 F:1306-1313/Region: nucleotide-binding motif A (P-loop)

Query Match 27.18; Score 1983.5; DB 1; Length 1515;
 Best Local Similarity 33.18; Pred. No. 6,7e-125;
 Matches 472; Conservative 268; Mismatches 492; Indels 195; Gaps 27;

```

QY 77 PKGHHGLSLAKRIRPTCKHOFVNDAGLSCMTFMSLSLARVARKKGLSMEDVSL 136
DB 196 PHOHITLIRRRK-----NPYDSANIFSRITFSMSGLMTGYEK YLVADLTKL 246
QY 137 SKHESSDYNCRRLERLMOEELNEVGPDAALRRVWIFCRT--RLILSYCLMITOLAG 193
DB 247 PRNSESSELSOKLEKMNENELKOKSNPSLS-----MAICRFGSKMLLAFFKAHIDVLA 301
QY 194 FSGP--AFWKHLEET-----QATESMLQYSLLYLGLL-----TE 229
DB 302 FTQQLRLIKFTYTDYNSERODHSSLOGFENNHPQLPIVRGFLLAFAFLVGTQTS 361
QY 230 IVRSMSLALFWALNYRTGVRLGAILTMAFKILKLNKE--KSLGELINCSGDCGM 287
DB 362 VLHOYELNV-----FNTGMKIKSALITALIYOKSLVLSNBSGLSTGDIYMLAVDYOKL 416
QY 288 PEAAVGSLLAGPVAAILGMIVVILGPTGFGSAVFILFYPAMMPASRLTAYFRKC 347
DB 417 QDLTQMLNLWISGFPQIILICLYSLKLLGNSMWGVILVLMPLNFMIRIQRKLOKSQ 476
QY 348 VAATDEVOKMNEVLYIKFKIYKAWYKAFSOSQVQIR-EEERILKAGFGOSITGVA 406
DB 477 MKYDETRVISELNLNKKISKLTYAMEKPYREKLEEVARNKELKNLTKLCYMAVTSPQF 536
QY 407 PIYVVASVTVFSVHM-TLGEDLTAQAFTVTVFNSMTFALKVTPSVKSLSEAVAYD 465
DB 537 NIYVELVSCCTFAFVYTERDALTTDLVFPALTLFNLSPMLTIIPWLSFIASVIG 596
QY 466 RFLSFLMEVEYHMKNNRPASPHIKIEKKNATLAMDSSHSSIQNSPKLTPKKKKRKRASRG 525
DB 597 RLFFEFNEEL-----OPDSVQRLPKY--KNIGVAINIG 629
QY 526 KKEKRVQLRTEHOAVLAEOKGHLLDSDERSPEEKGKHIHGLRLQTLTSLDEI 585
DB 630 DDAFTLMQRREPKYV-----ALKNINPQA 653
QY 586 QEGKLVIGSGVSGSKTSLISAILGQMTLLEGSIAISGTFAYVAQAQMLNATLRDLIF 645
DB 654 KKGMLTCTIVGKVGSGTALLSCMLGDLFRVKGFAIVHGSVAIVYQVPMINMGTVKENILF 713
QY 646 GKEVDERYNSVNSCCRLDIALTPSSDLTEIGERGANLSGGQRQRISTARALYSRST 705
DB 714 GHRDADFYEKTIKACALTTIDLATLMDGDTLVGEGKISLSGGQKARISLARAYARADT 773
QY 706 YILDDPLSALDAHNGNIIFNSAIRKH--LKSIVYLVTHOLOVYVDDDEVIFMKEGCIT 763
DB 774 YLDDPLAADAENHARLIEHVLGPNLHRTKYATLANKYALSADSLADSLANGELTQ 833
QY 764 RGHHEEL-ANLNGDYATIFENLLIGETPVEINSKETSQSKQSDKGPRTGSIKKEKA 822
DB 834 QGVYDEITTKDADSPMLKLNNY-----GKKNNG--KSNFEGDSSESVRESS 878
QY 823 VKPEEGQLVQ-----EKGCGGSVPMS 844
DB 879 I-PVEGLELDLOKINDLDFGNSDAISLRASDATLGIDGDDENIAKREHREGQKWKVN 937
QY 845 VGYVYIOAGGPIAFIVYIMALFMLNVSTAFSTWMLSYWIKQSGNGTTVTNGNTSVDSD 904
DB 938 IYLEYAKAC-NPKSVCYFIIFIVISMFLSVMGVNLKHW-----SEVNSR 981
QY 905 MKDNPHQVYASIV--ALSMAMVLLIKAIRGVVYVKGTLRASSRLHDEFLRILRSRPMK 961
DB 982 YGSNPNARALATAYFALGIGSALATLTIQTI-VLWVCTHASKYLLHNTMNSVLRAPMT 1039

```

```

QY 962 FEDTTPGRLINRFSKMDDEVDRILPFOAEMFIQNVILVEFCVMAGVPEWLVAVGPI 1021
DB 1040 FFETTPGRLINRFSNNDIYKDALGRFSSQFVYNAVKVFTPIVATWQFIIFIIPL 1099
QY 1022 VILFSLHIVSRVLIREKRLNDTOSPFSLSHITSIOGLATIHANYKGOEFLRYOGL 1081
DB 1100 SVFTIYQOYYLRFSRELRRDSTIRSPISYHFOETLGGTLVGVYSOQKRFSHINCRI 1159
QY 1082 DDNOAPPELFYCAMRLAVRLDLISALITTTGLMIY-LMHGOIPAPYAGLASYAVOL 1139
DB 1160 DNMSAFPSINARMLAVRELIGSITILGAATLSVFRLKOGFLTGMVGLSLSYALQI 1219
QY 1140 TGLFOFTVRLASEPAREFTSVERNHYIKTSLAPARIKAPASPDMPQDEVTENAE 1199
DB 1220 TQTLNWTIVRMTEVEVETNIVSVERIKET-ADLKSSEAPLVEGHRPPKMPSPQDIFKNYS 1278
QY 1200 MRYRENPLVLKKSFTIKREKIGIVGRGSGSSISGMALFRLEVELSGGCIKIDGVRIS 1259
DB 1279 TRYPELDVLKHNINHIKPEKKGIVGRGAGKSSLTALFRMIESEGNVIDNAIN 1338
QY 1260 DIGLADLRSKLSIIPQEVLFSGTVRSNLDPEFNQYTEDQIDALEPTHMKECIAQLPLK 1318
DB 1339 EIGLYLRHKLSIIPQDSQVEGVRENDIPNGYTDPAIRALELSHKHVLMSNDG 1398
QY 1319 LSEVMEKDNFVSGERQLCTIARALLRCKILLDERTAMDETLLIOETIRAFAD 1378
DB 1399 LDAOITREGCMVSGORQLCLARALVPSKTIIVDEXTAAVDDETIRKVOETIRTKFD 1458
QY 1379 CTMLTIARLHTVLGSDRIWLAQOVVEFPTPSVLLSNDSSRYAM 1425
DB 1459 RTIITLIRLNTIMDSRIIVLDGKVAEPDSQGLSDNKSLEYSL 1505

```

RESULT 11
 JE0336
 canalicular multispecific organic anion transporter - human
 C:Species: Homo sapiens (man)
 C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 04-Jun-1999
 C:Accession: JE0336
 R:Uchiumi, T.; Hinoshita, E.; Haga, S.; Nakamura, T.; Tanaka, T.; Toh, S.; Furukawa, Biochem. Biophys. Res. Commun. 252, 103-110, 1998
 A:Title: Isolation of a novel human canalicular multispecific organic anion transporter.
 A:Reference number: JE0336; MUID:99032812
 A:Accession: JE0336
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1527 <CDS>
 A:Cross-references: GB:AF083552
 C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homolo
 C:Keywords: ATP
 F:1306-1499/Domain: ATP-binding cassette homology <ABC2>

Query Match 26.68; Score 1941; DB 2; Length 1527;
 Best Local Similarity 32.28; Pred. No. 4.9e-122;
 Matches 463; Conservative 268; Mismatches 511; Indels 196; Gaps 22;

```

QY 85 LSAI-----KPIRTCK-----HQHPVDNAGLSCMTFMSLSLARVARKKGLSMEDV 133
DB 182 LSAIILACFRREKPPFAKKNVDPNPYPETSAGFLSRLFWMFTKMAIYGRH-PLKEDL 240
QY 134 WLSKHESSDYNCRRLERLMOEELNEVGPDAAS-----L 167
DB 241 WLSKEEDRSQVYQQLLEAWRKQKQTAHRAASAAPKKNASGEDEVLLGARPRRKPSPFL 300
QY 168 RRVWICRRRLIISYCLMITOLAGSGPAFMVKHLELYQATENSMLQYSLLYLGLLL 227
DB 301 KALATP-GSSEFLISACFKLIQDLSTINPQ-LISILIRFSNPMGSGWGWGFLVAGLMFL 358
QY 228 TEIYRMSLALFWALNYRTGVRLGAILTMAFKILKLN-IRKES-LGELINCSNDGQ 285
DB 359 CSMMQSLLOHYHYITVYGVKFRGTGIMGYIRKALVITNSVKNASIVGELVNLMSVDAQ 418

```

| | | | |
|----|------|--|------|
| Qy | 286 | RMEFAAVSGSLLAGPVALIGIYVWIIILGPRGIFGSVFLIFPAPMAMFASRLATYRR | 345 |
| Db | 419 | REMDLAPFLNLSAPLOIILATIFYLMOQNLGPGSVLAGVAFWLLIPLNGAVAKMRAOV | 478 |
| Qy | 346 | KCVATADERVOKMNEVLTJIKFIKMYAMKAFOSVOKFREERIRILEKAGYFOSITGV | 405 |
| Db | 479 | KOMLKORSRIKMSIELINGIKVILKVLAMBSPLKOVDEGRGDELQRLRAALHTHTTFT | 538 |
| Qy | 406 | ---APIYVJLASVYTSVAMTKLGFDLTAAOAFVTYVENSMTFALKVYTPESVKSISEASV | 462 |
| Db | 539 | WMSBPFLVTITITIMVY-VYVDPNNVLDIAEKAFYSVSLFINILPLMLPOLISNTLOASV | 597 |
| Qy | 463 | AVDSFKSLFIMEEV--HMINKRASPARIKEMKANLADSSHSIIONSPKLTPMKMKD | 520 |
| Db | 598 | SLAKIOQFLSOEBELDPOSVYRKTIISGCIYATIIISGFTYAO----- | 639 |
| Qy | 521 | RASRKKERYKQLORTBHOAVLAEOKGHLLDSDERPSPREEGKHILGHLRLORTHS | 580 |
| Db | 640 | -----LPPTLHS | 646 |
| Qy | 581 | IDLEIOGKLVIGOGSVGSKTSLISAIIGOMTLLEGSLAISGTFAYVQAOQWIIATLR | 640 |
| Db | 647 | LDIOVPAGLAAVAVGPGCKSSVLSALJGEMKIEKGKVMHGSVAAYVPOQMIONCQLQ | 706 |
| Qy | 641 | DNILFKEYDEERYENSVJLNSCCRLPDLALIPSSDLIEIERBANTSGGORISITARALY | 700 |
| Db | 707 | ENVLFGRALNPKRYOQLEACALLADDEMPPGDOYEIEKGINISGGORQVSLARAVY | 766 |
| Qy | 701 | SDRSIYIILDDPLSALDAHVONHIFNSAIRKH--LASKYVLYPTHOLOLYVDCDEVYFKE | 758 |
| Db | 767 | SDADIFILDDPLSLVSDSHVAKHIFDHVIGEGVLAKTIVLTHGISLPTQDFTIYIAD | 826 |
| Qy | 759 | GCITERGTHEELMLNDYA-----TIFNN | 783 |
| Db | 827 | GQVSEMGYPRIALLQNRGSPFANPLCNAPDEBDGHLDESWTALBGAEDKALLIEDTLSNH | 886 |
| Qy | 784 | LLIGETPPV-----BINSKKTSGSQKSDQKGRPTSGIKKRAKPEEGOL | 830 |
| Db | 887 | TDLINDNPVIYVVOKOPRQSLALSJSDGEGGPRVRRHIGP--SEKQVTEAKADAL | 943 |
| Qy | 831 | VOLEEKGOGSPVMSVYVYIOAAGGFLAVLYMALMLVNGSAF---STMWLSWINGO | 887 |
| Db | 944 | TOEKRAALIGVEISVFPDYAKAVG---LCTTATICLLVYGSSAAAGVNLMSW---- | 995 |
| Qy | 888 | SGNTVTVRGNETSVSDSMKPNPMHOYASIALYSMAVMMLIKAIRGVVFGKLTASSRL | 947 |
| Db | 996 | -----FMDAADSROUNTSLR--LGYVALGILGOFIVMLAMMAAGIOQAVL | 1044 |
| Qy | 948 | HDELFRILNSPKAFPTTPTTGRIILNRFESKDMDDEVYRLPFOAEFIQVNLVFFCVGKI | 1007 |
| Db | 1045 | HQALLHNKIRSPQSEFDTTSGRIILNCFSDIYVDEVAVLPIMLNLFENAISTLVYI | 1104 |
| Qy | 1008 | AGVPMFLVANGPVLPIVLFVSILVHSRVLJLEKRLDNITIOSPFLSHITSIGLTIHAY | 1067 |
| Db | 1105 | MASRPLETVVILPLAVLYTLVQRFAYATISQLRKLESVRSRPIYSHIFSTYTGASVITAY | 1166 |
| Qy | 1068 | NKGQELHRYOEILLDNOAPFLFETCARMRLAVRLDISALITTTGMLVIMHQIPPA | 1127 |
| Db | 1165 | NRSDPELISPTKVDANQORSQCYIISNRLSIGVEFVNCVULFALPAVIGRSSIMBG | 1224 |
| Qy | 1128 | YAGLAISYAVOLGSLPOFYVRLASEPTEARTSYERLNHYIKTLSLEAPARIKNKAPSDW | 1187 |
| Db | 1225 | LVGSLSVSYLQVTFALMMIMRMSDLESNTIVAAERKEYSKT-ETFAPVVYVGSRRPGW | 1283 |
| Qy | 1188 | POGEVYFENAKMERENLPLVLKVVYSFTIKPEKXIGIYGRGSGKSSLGMLFRLVEIS | 1247 |
| Db | 1284 | PPRGVEVERNYSVYRGRGLDVLRLDSLHNHGEKXIGIYGRGAGKSSMTLCFLFILLEA | 1344 |
| Qy | 1248 | GGCIRKIDGVRISDIGLADLSKSIITPOEVVLFSGTVRSNLDPPFNOYTEDQIWDALERTH | 1307 |
| Db | 1344 | KGELIRIGELNVAJDGLADVSOQLTIIPDPILESFGLRNLNDPPFGSYSESDITWMLLEISH | 1403 |

[illegible]


```

QY 773 -----LNGDYATIFNNLLGSETPPE-----IN 795
Db 890 RERREAEASADDEDDENSEPBGIMIGDSDFEYDDDDVMA--SPITIDHVLGTSHMSTVSGIIN 948
QY 796 SKKETSGOKSQDQKPKTKSIKKEKAVPEEGOLVLEEEKGCGSVPMVGYGYYIOAGG 855
Db 949 RRRISTSTHKQRRLSTSTSHSHITSASTQRLGTGRVETGRKMTYYKKYKFCAMQM 1008
QY 856 PLATFVLMALFMLNAGSTAFS---TWMLSYWKQSGNTTGTGNETSVSDSKMDPNHQ 912
Db 1009 STAVLEVLGM-----TSTSTFSGMRNMLTFDW-----SNDNAAGSGSNTT-----GQPIA 1053
QY 913 YVASIYALSMAMMLIKAKRGVFPVKGTIRASSRLHDELFRRLSPKMPFDTTPGRLL 972
Db 1054 IRLGAYAGLGFSEIILLFTGMLISLLYGVSASRNHAPMLRMLFRVPMAYDTPTPGRLL 1113
QY 973 NRRFSKDMDEVDRLPFOAMFIQNVILVFCVGMAGVPMFLVAGPVLVLEFSVLHYS 1032
Db 1114 NRIQGDIEVDVLPNNQVFOACLOVSTLIIMISTPVGIVYIPLSVATLMRMRY 1173
QY 1033 RVLRELKRLDNITOSPFLSHITSSIOGLATTHAYNKGOEFLHRYQELLDNQAPFLEF 1092
Db 1174 IATSRQLKRESITRSPISYSHLSSESIOGSATIRAYHLVDRFCKLSETKYDSHVQCRYLNY 1233
QY 1093 CAMRMLAVRLDIALITTTGIMIVLMHGQIPRANAGLAISAVOGLGLOFTVRLASE 1152
Db 1234 VANRMLSVLEFIQNCIVLSALFPALPRTTTTSGVIGSVSALNITVNLNVAQDITK 1293
QY 1153 TEARFTVERINHXYIKTSLLEAPARIK---KAPSPDMPQEGEVTEPENAEMRYREMLPV 1209
Db 1294 LRTNIVSVRVEYAEET---ETEAEMKSEPGKEPQGNPSEGIYVANNYSARRRPLNY 1350
QY 1210 LKKVSTTKPKKIGIVGTGSGKSLGMAFLRVELSGGCIKIDGVRIISDGLADLRK 1269
Db 1351 VOLNWEIKPHEKVGIVGTGAGKSSVTSLFRILTEAAGQIVYDGINLAETGLHDLRSN 1410
QY 1270 LSTIOEPVLFSGTVASNDPNOYTEDQIMDALERTHKECTAQLPLKLESEVMNGDN 1329
Db 1411 LRTIPDPVLESGTILFENLDPNNHSDGIMIKTLEMANIKERTAHNEQNTYITGGDN 1470
QY 1330 FSVGEROLLICARALLRHCKIILDEATAMDTEDLLIQETIREAFADCTMLTIAHRL 1389
Db 1471 ISVGRQQLACARALLRKRLVILDEATAVYSDTALQKTRREFANATVLTIAHRLN 1530
QY 1390 TVLGSDRIMVLAOGVVEEDTPSVLLSNDSSRYAM 1425
Db 1531 TMDYDRITIVLNDGKVEEDSPANLISNNSEFSYM 1566

RESULT 14
T18059
ABC transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein F26K9.130
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 21-Jul-2000
R:Accession: T18059
R:Biocheck, H.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quettier, F.; Salanoubé
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224465
A:Accession: T18059
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1539 <BL0>
A:Cross-References: EMBL:AL162651
A:Experimental source: cultivar Columbia; BAC clone F26K9
C:Genetics:
A:Map position: 3
A:Introns: 804/3; 833/3; 1078/3; 1133/3; 1232/1; 1405/3; 1427/1; 1507/1
C:Note: F26K9.130
C:Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology

```

```

Query Match      25.8%; Score 1882; DB 2; Length 1539;
Best Local Similarity 32.2%; Pred. No. 4.7e-118;
Matches 454; Conservative 277; Mismatches 485; Indels 194; Gaps 30;

QY 102 DNAGLFSMTF-----SWLSLARVAHKKGELSMEDVSLSKHSSDYNCRLERLW-Q 154
Db 245 ENVSIVASASFSISKTPWLMNPNLRKGY--KSPLNLDQVPTLSPHRAKLTATLESKWPK 303
QY 155 EELNEVGPDAASLRVWYFICRRLILSTVCLMTIQLAQSOPAFVVKLLLETQATBSN 214
Db 304 PONSNNPVRTLLIRCFWEIAFTVAIAIRLSVI---YVGPV-LIOSFVFTSGKRS 358
QY 215 LQYSLLVYGLLTETELVRSMLATFWALNY---RQVRLGAILVMAFKILIKNIKER 271
Db 359 PEOGYTVLILILAKFEVLS---THQFNNSOKLGLMLRSTLITLTKKGLKLTGSAAQ 415
QY 272 S--LGLINICNSDQGR---MFEAAVGSLLAGFPVAIILGIVNVIILGPV-----G 319
Db 416 NNGQGVAVYMAVDQAQLDDMLQLHAI--WMLPLQVAIAVLLVNT--LGPVTVTVIG 471
QY 320 FLGSAVFLIFPAMFASRLTAFRRKCYAADDERVQKANEVLYTKFKTKMAWKAQSO 379
Db 472 LGIGIFVFIIL-----LGTRRNNRYOPSLMNNRDSRRKATNEMLNMYRIKFOAMEDHNE 525
QY 380 SVQKIREERRLIEKAGYQOSITGVGAPIVVYVYVFSVHMTLGEDLLAQAQFTVVTV 439
Db 526 RILKPREMEFGWLKSKFLYSIAGNIYLMSTPVLSALITTTAVFLQVLDAGVFTTTTI 585
QY 440 FNSMTPALKVTPFSYKSLSEASVADRFKSLFLMEVH---MINKNPASPHIKIEKNAT 496
Db 586 FKLIQPIRITFPQSMISLSQAMISGLRDAYMMRSRLSEFTVRSRSGCGGNVAVEIKDS 645
QY 497 LAMDSHSSITQNSPKLTPMKKDKRASRGKKEVROLQTRHOAVYLABOKGHILLDSER 556
Db 646 FSRD-----DEDE 654
QY 557 PSPREEEGKHILHRLRLQRTLSIDLEIOEGLVIGCSVSGSKTSLAISILIQMTLE 616
Db 655 PA-----IENIFEVKKGELAIYGVTSKGKSSLASVYGEHKLIS 695
QY 617 GSIAISGTFAYAAQAMILNATLRDLNLIIFGEKEYDEERYNSVNSCLRPDLALPSSDLT 676
Db 696 GKRVCGTAAVYAQTSWIONGVQDNLIFGLPNNRSGKYNEVLKCCLEKDMQMEGGDT 755
QY 677 EIGERANISGGQRORISLARALYSRSTIYIDDPISALDAHNGNIIFSAIRKHLKST 736
Db 756 EIGERGINLSGGKORIOLARAYQESDYLLDDVFSADVAHGSDFIKKCVAGALKRGT 815
QY 737 VLEFTHQLQYLVDCDEVIFMKEGCIERGEHEELMNLNGYATIFNNLLGSETPP--VET 794
Db 816 ILLVTHQVDFLHNVDKILVNRDGMATVQSGKIDELVSSGLDFGL--VAAHETSMELVER 872
QY 795 NSKKESTGS-----QKKSODKGRK--TGSIKKEKAVK----- 824
Db 873 GSAASATAANVPMAAPITQRSISTIESPROKSPKVRHRTSMESPRVLRITSMESPRLELN 932
QY 825 -----PEEG-QLVQLEEEKGCGSVPMVGYGYYIOAGGRLATFVLMALFMLNAG 871
Db 933 DESIKSFILSNIPEDSGRLIKEREVEGVSFQVYLYLSTEALGWMGMLILVFFSVAMQA 992
QY 872 SFAFSYMWLSYWKQSGNTTGTGNETSVSDSKMDPNHQYVASIYALSMAMMLIKAI 931
Db 993 SLMASDYWLAY-----ETSAKNVSDAYV--FIRYVIAANSIYLVCL 1035
QY 932 RGVVFPKGTLRASSRLHDELFRRLRSPKMPFDTTPGRILNRRFSKDMDEVDRLPQ-- 989
Db 1036 RAYVYTHLGKTAQIFPKQILNLSLVHAPMSFPTPSGRILSRASDQNVNVDIFIPMIG 1095
QY 990 --AEMFIQNVILVFCVGMAGVPMFLVAGPVLVLEFSVLHYSVNLIREKRLDNITQ 1047
Db 1096 LVATMT--TLLSLTIVTCQYAMPVTFEILPLGLWLNTRYKYIYLAAS--RELRLDITIK 1151
QY 1048 SPLSHITSSIOGLATTHAYNKGOEFLHRYQELLDNQAPFLETCAMRMLAVRLDLIS 1107

```


[illegible]

Search completed: November 9, 2001, 17:48:05
Job time: 6553 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2001, 17:48:02 ; Search time 78.74 Seconds
(without alignments)
625.160 Million cell updates/sec

Title: US-09-528-031-2
Perfect score: 7308
Sequence: 1 MKDIDIGKEYIIPSGYRSV.....DSRFYAMFAAENKVAVKG 1437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------|---------------------|
| 1 | 7297 | 99.8 | 1437 | MRP5_HUMAN | O15440 homo sapien |
| 2 | 6923.5 | 94.7 | 1436 | MRP5_RAT | O9qym0 rattus norv |
| 3 | 6855.5 | 93.8 | 1436 | MRP5_MOUSE | O9jix5 mus musculu |
| 4 | 2100.5 | 28.7 | 1541 | MRP2_RAT | O63120 rattus norv |
| 5 | 2082 | 28.5 | 1531 | MRP1_HUMAN | P33527 homo sapien |
| 6 | 2073 | 28.4 | 1545 | MRP2_HUMAN | O92887 homo sapien |
| 7 | 2053.5 | 28.1 | 1325 | MRP4_HUMAN | O15839 homo sapien |
| 8 | 2033.5 | 27.8 | 1564 | MRP2_RABIT | O28689 oryctolagus |
| 9 | 1983.5 | 27.1 | 1515 | YCF1_YEAST | P39109 saccharomyc |
| 10 | 1960 | 26.8 | 1522 | MRP3_RAT | O88563 rattus norv |
| 11 | 1945 | 26.6 | 1527 | MRP3_HUMAN | O15438 homo sapien |
| 12 | 1874.5 | 23.6 | 1478 | YAMB_SCHPO | O10185 schizosacch |
| 13 | 1759 | 23.1 | 1580 | ACC8_HUMAN | O09428 homo sapien |
| 14 | 1749 | 23.9 | 1592 | YHDS_YEAST | P38735 saccharomyc |
| 15 | 1742.5 | 23.8 | 1581 | ACC8_CRICR | O09427 cricetus cr |
| 16 | 1733.5 | 23.7 | 1559 | BPT1_YEAST | P14772 saccharomyc |
| 17 | 1732 | 23.7 | 1580 | ACC8_RAT | O09429 rattus norv |
| 18 | 1727 | 23.6 | 1477 | YOR1_YEAST | P53049 saccharomyc |
| 19 | 1700.5 | 23.3 | 1503 | MRP6_HUMAN | O95555 homo sapien |
| 20 | 1665 | 22.8 | 1502 | MRP6_RAT | O88266 rattus norv |
| 21 | 1649.5 | 22.6 | 1601 | YPR1_YEAST | P32386 saccharomyc |
| 22 | 1532 | 20.8 | 1548 | MDR1_YEAST | P21441 leishmania |
| 23 | 1451 | 19.9 | 1492 | CFTR_SQUAC | P26362 squallus aca |
| 24 | 1408 | 19.3 | 1480 | CFTR_HUMAN | P13369 homo sapien |
| 25 | 1405.5 | 19.2 | 1558 | YK03_YEAST | P36028 saccharomyc |
| 26 | 1381 | 18.9 | 1481 | CFTR_SHEEP | O00555 ovis aries |
| 27 | 1378 | 18.7 | 1476 | CFTR_MOUSE | P26361 mus musculu |
| 28 | 1369 | 18.7 | 1481 | CFTR_BOVIN | P35071 bos taurus |
| 29 | 1365.5 | 18.7 | 1485 | CFTR_XENTA | P26363 xenopus lae |
| 30 | 1334.9 | 18.5 | 1427 | ABCL_SCHPO | O92337 schizosacch |
| 31 | 1329.5 | 18.2 | 1450 | CFTR_RABIT | O00554 oryctolagus |
| 32 | 785.5 | 10.7 | 1276 | MDR2_CRIGR | P21449 cricetus cr |
| 33 | 774.5 | 10.6 | 1321 | MDR1_CAEBL | P34712 caenorhabdi |

| | | | | | |
|----|-------|------|------|------------|--------------------|
| 34 | 767.5 | 10.5 | 1280 | MDR1_HUMAN | P08183 homo sapien |
| 35 | 766.5 | 10.5 | 1276 | MDR1_CRIGR | P21448 cricetus cr |
| 36 | 764.5 | 10.5 | 1276 | MDR1_MOUSE | P06795 mus musculu |
| 37 | 764.5 | 10.5 | 1281 | MDR3_CRIGR | P23174 cricetus cr |
| 38 | 759.5 | 10.4 | 1278 | MDR2_RAT | O08201 rattus norv |
| 39 | 754.5 | 10.3 | 1277 | MDR1_RAT | P43245 rattus norv |
| 40 | 752 | 10.3 | 1276 | MDR2_MOUSE | P21440 mus musculu |
| 41 | 750.5 | 10.3 | 1279 | MDR3_HUMAN | P21439 homo sapien |
| 42 | 740.5 | 10.1 | 1302 | MDR4_DROME | O00449 drosophila |
| 43 | 735.5 | 10.1 | 1276 | MDR3_MOUSE | P21447 mus musculu |
| 44 | 656.5 | 9.0 | 1362 | PNM1_SCHPO | P36619 schizosacch |
| 45 | 638 | 8.7 | 1290 | STE6_YEAST | P12866 saccharomyc |

ALIGNMENTS

RESULT 1
ID MRP5_HUMAN STANDARD; PRT; 1437 AA.
AC O15440: Q9UOC3; Q9UNP5; Q9UN85; O14517;
DT 15-JUL-1998 (Rel. 36, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 5 (ABC TRANSPORTER MOAT-C)
DE (PABCL1) (SMRP).
GN ABCS OR MRP5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99043202; PubMed=9827529;
RA Belinsky M.G., Bain L.J., Balsara B.B., Testa J.R., Krueh G.D.;
RT "Characterization of MOAT-C and MOAT-D, new members of the MRP/cMOAT
RT subfamily of transporter proteins.";
RL J. Natl. Cancer Inst. 90:1735-1741(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99367488; PubMed=10438534;
RA McAlleer M.A., Breen M.A., White N.L., Matthews N.;
RT "PABCL1 (also known as MOAT-C and MRP5), a member of the ABC family of
RT proteins, has anion transporter activity but does not confer multidrug
RT resistance when overexpressed in human embryonic kidney 293 cells.";
RL J. Biol. Chem. 274:23541-23548(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20184734; PubMed=10721709;
RA Suzuki T., Sasaki H., Kuh H.J., Agui M., Tatsumi Y., Tanabe S.,
RT Terada M., Saijo N., Nishio K.;
RT "Detailed structural analysis on both human MRP5 and mouse mrp5
RT transcripts.";
RL Gene 242:167-173(2000).
RN [4]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Brain, and Ovarian carcinoma;
RX MEDLINE=20319051; PubMed=10840050;
RA Wijnholds J., Mol C.A.A.M., van Deemter L., de Haas M., Scheffer G.L.,
RA Baas F., Beijnen J.H., Schepers R.J., Hulse S., De Clercq E.,
RA Balzarini J., Borst P.;
RT "Multidrug-resistance protein 5 is a multispecific organic anion
RT transporter able to transport nucleotide analogs.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:7476-7481(2000).
RN [5]
RP SEQUENCE OF 492-1437 FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97472289; PubMed=9325169;
RA Suzuki T., Nishio K., Sasaki H., Kurokawa H., Saito-Ohara F.,
RA Ikeuch T., Tanabe S., Terada M., Saijo N.;
RT "CDN cloning of a short type of multidrug resistance protein
RT homologue, SMRP, from a human lung cancer cell line.";

RL Biochem. Biophys. Res. Commun. 238:790-794(1997).
RN [6]
RP SEQUENCE OF 1216-1437 FROM N.A.
RC TISSUE-Brain;
RX MEDLINE-97413640; PubMed-9270026;
RA Kool M., de Haas M., Scheffer G.L., Scheper R.J., van Eijk M.J.,
RA Juljn J.A., Baas F., Borst P.;
RT "Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5,
RT homologues of the multidrug resistance-associated protein gene
RL (MRP1), in human cancer cell lines";
RL Cancer Res. 57:3537-3547(1997).
CC -1- FUNCTION: ACTS AS A MULTISPECIFIC ORGANIC ANION PUMP WHICH CAN
CC TRANSPORT NUCLEOTIDE ANALOGS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF104942; AAD04169.1; -
DR EMBL; AF146074; AAD37716.1; -
DR EMBL; AB019002; BAA76608.1; -
DR EMBL; AB005659; BAA22887.1; -
DR EMBL; U83661; AAB71758.2; -
DR InterPro; IPR001140; -
DR InterPro; IPR001617; -
DR Pfam; PF00664; ABC_membrane; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 219 239 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
FT TRANSMEM 608 628 POTENTIAL.
FT TRANSMEM 848 868 POTENTIAL.
FT TRANSMEM 917 937 POTENTIAL.
FT TRANSMEM 997 1017 POTENTIAL.
FT TRANSMEM 1018 1038 POTENTIAL.
FT TRANSMEM 1104 1124 POTENTIAL.
FT TRANSMEM 1127 1147 POTENTIAL.
FT NP_BIND 595 602 ATP (POTENTIAL).
FT NP_BIND 1227 1234 ATP (POTENTIAL).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 1329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1417 1417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 176 176 R -> P (IN REF. 3).
FT CONFLICT 400 400 S -> G (IN REF. 1).
FT CONFLICT 581 581 I -> V (IN REF. 2).
FT CONFLICT 1383 1383 T -> N (IN REF. 3 AND 5).
SQ SEQUENCE 1437 AA; 160659 MW; 00558076B3BB4C00 CRC64;

QY 61 LDASHNSQLRLDEEHPRKXHHGSLALKPIRTTCKHOPVDNAGLSQMTFMSWLSLAR 120
DB 61 LDASHNSQLRLDEEHPRKXHHGSLALKPIRTTCKHOPVDNAGLSQMTFMSWLSLAR 120
QY 121 VAHKGELSMEDVWSLSHSSDVNCRRLERLMOELNEVPDASLRVYVWICRRLI 180
DB 121 VAHKGELSMEDVWSLSHSSDVNCRRLERLMOELNEVPDASLRVYVWICRRLI 180
QY 181 LSIYGLMTQLAGFSGPFWKHLLEYQATESNLQYSLLVGLLLEIYRMSLALTW 240
DB 181 LSIYGLMTQLAGFSGPFWKHLLEYQATESNLQYSLLVGLLLEIYRMSLALTW 240
QY 241 ALNRYTVRLGALITPAAPKILKIKIKESLGLINICSDGQRMFEAAVAGSLAGG 300
DB 241 ALNRYTVRLGALITPAAPKILKIKIKESLGLINICSDGQRMFEAAVAGSLAGG 300
QY 301 PVVALIGMIVNYILGPGFGLSANFELFYAPAMPASRLTYFRKCYAANDERQKNE 360
DB 301 PVVALIGMIVNYILGPGFGLSANFELFYAPAMPASRLTYFRKCYAANDERQKNE 360
QY 361 VLTYYKFKMTAWKAFQSQYOKIREERRLIEKAGYFQSTVGVAPIVVYIASVYFVS 420
DB 361 VLTYYKFKMTAWKAFQSQYOKIREERRLIEKAGYFQSTVGVAPIVVYIASVYFVS 420
QY 421 HMTLGFDLTAQAFTVTVVNSMTFALKVTPSVKSLSEASVANDRFSLTMEVHIK 480
DB 421 HMTLGFDLTAQAFTVTVVNSMTFALKVTPSVKSLSEASVANDRFSLTMEVHIK 480
QY 481 NKPAAPHIKIEKKNTLMDSSHSIONSPLTPKMKDKRASRCKKRYKQLOTEHQ 540
DB 481 NKPAAPHIKIEKKNTLMDSSHSIONSPLTPKMKDKRASRCKKRYKQLOTEHQ 540
QY 541 VLAEOGHLLDSDERPSPREEEGKHILGHLRLORTLSIDLIEQEKLVGICGSVSG 600
DB 541 VLAEOGHLLDSDERPSPREEEGKHILGHLRLORTLSIDLIEQEKLVGICGSVSG 600
QY 601 KTSLSAILGQTLLEGSAISGTFAYVAQQAAMILNATLRDMLIEGKYDEERYSVNS 660
DB 601 KTSLSAILGQTLLEGSAISGTFAYVAQQAAMILNATLRDMLIEGKYDEERYSVNS 660
QY 661 CCLRPDLAIPSSDLTEIGERGANLSCGORISLARALSPRSYIIDDPLSALDAHV 720
DB 661 CCLRPDLAIPSSDLTEIGERGANLSCGORISLARALSPRSYIIDDPLSALDAHV 720
QY 721 NHIFNSAIRKHLKSTVLFVTHQLQYLVDCEVIEFMKECCTERGTHELMNLDVATI 780
DB 721 NHIFNSAIRKHLKSTVLFVTHQLQYLVDCEVIEFMKECCTERGTHELMNLDVATI 780
QY 781 FNNLLGEPPEIETSKETSGSKKSDKPKGTGSIKKKAVKPEEGOLVLEKGGGS 840
DB 781 FNNLLGEPPEIETSKETSGSKKSDKPKGTGSIKKKAVKPEEGOLVLEKGGGS 840
QY 841 VPMASYGVYIOAGGPLAFVLMALFMLVNSTAFSTWMLSWIKSGSNTVTGNETS 900
DB 841 VPMASYGVYIOAGGPLAFVLMALFMLVNSTAFSTWMLSWIKSGSNTVTGNETS 900
QY 901 VDSMKNDPNHMYASTIALSMAVMLILKAIKGVVFKGTLRASSRLHDELFRILRSPM 960
DB 901 VDSMKNDPNHMYASTIALSMAVMLILKAIKGVVFKGTLRASSRLHDELFRILRSPM 960
QY 961 KFFDTTPGRIILNRSKMDDEVRLPQAEFIONVILVFPCVMINGVFWFVLAAGP 1020
DB 961 KFFDTTPGRIILNRSKMDDEVRLPQAEFIONVILVFPCVMINGVFWFVLAAGP 1020
QY 1021 LVILPSVLAHSRVILRELKRLDNTOSPFSLHSSIOGLATIHAYNKGOFELRLYOEL 1080
DB 1021 LVILPSVLAHSRVILRELKRLDNTOSPFSLHSSIOGLATIHAYNKGOFELRLYOEL 1080
QY 1081 LDDNAPPELFTCAMRWLAIVRLDLISALITTTGLMIVLMHGOIPPAYAGLAISYAVOLT 1140
DB 1081 LDDNAPPELFTCAMRWLAIVRLDLISALITTTGLMIVLMHGOIPPAYAGLAISYAVOLT 1140

QY 1141 GLEFOTVRLASEETAEPTVERINHYIKTSLSEAPARIKKKAPSPMPQGEVTEFENAEK 1200
 DB 1141 GLEFOTVRLASEETAEPTVERINHYIKTSLSEAPARIKKKAPSPMPQGEVTEFENAEK 1200
 QY 1201 RRENPLVLKRVSTIKERKIGIVGTGSGKSGMALFRLVELSGCCKIDGVRISD 1260
 DB 1201 RRENPLVLKRVSTIKERKIGIVGTGSGKSGMALFRLVELSGCCKIDGVRISD 1260
 QY 1261 IGLADRSKLSITIPQEPVLFSGTVRSNLDPEFNOYTEDQIWDALERTHMECIAQLPLKLE 1320
 DB 1261 IGLADRSKLSITIPQEPVLFSGTVRSNLDPEFNOYTEDQIWDALERTHMECIAQLPLKLE 1320
 QY 1321 SEVMENGDNFSGVEROLLIARALLRCKILLIDEATPAAMDTELDLIQETIREAFADCT 1380
 DB 1321 SEVMENGDNFSGVEROLLIARALLRCKILLIDEATPAAMDTELDLIQETIREAFADCT 1380
 QY 1381 MULTAHLHTVLGSDRIMVLAGOVVEFPTPSVLSNDSSREYAFMAAENKVAVGK 1437
 DB 1381 MULTAHLHTVLGSDRIMVLAGOVVEFPTPSVLSNDSSREYAFMAAENKVAVGK 1437
 RESULT 2
 MRP5_RAT STANDARD; PRT; 1436 AA.
 AC 090YMO;
 DT 01-OCT-2000 (rel. 40, Created)
 DT 01-OCT-2000 (rel. 40, Last sequence update)
 DT 01-OCT-2000 (rel. 40, Last annotation update)
 DE MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 5.
 GN MRP5.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
 RA Homma M., Suzuki H., Sugiyama Y.;
 RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: ACTS AS A MULTISPECIFIC ORGANIC ANION PUMP WHICH CAN
 TRANSPORT NUCLEOTIDE ANALOGS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 (ABC TRANSPORTERS). MRP SUBFAMILY
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB020209; BAA88897.1; -;
 DR InterPro: IPR001140; -;
 DR Pfam: PF00064; ABC_tran; 2.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 219 239 POTENTIAL.
 FT TRANSMEM 236 316 POTENTIAL.
 FT TRANSMEM 317 337 POTENTIAL.
 FT TRANSMEM 400 420 POTENTIAL.
 FT TRANSMEM 426 446 POTENTIAL.
 FT TRANSMEM 508 628 POTENTIAL.
 FT TRANSMEM 647 867 POTENTIAL.
 FT TRANSMEM 847 867 POTENTIAL.
 FT TRANSMEM 916 936 POTENTIAL.
 FT TRANSMEM 996 1016 POTENTIAL.
 FT TRANSMEM 1017 1037 POTENTIAL.
 FT TRANSMEM 1101 1121 POTENTIAL.
 FT TRANSMEM 1126 1146 POTENTIAL.

FT NP_BIND 595 602 ATP (POTENTIAL).
 FT NP_BIND 1226 1233 ATP (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 636 636 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 684 684 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 889 889 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 896 896 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1043 1043 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1328 1328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1436 AA; 160855 MW; 10FE53B800531598 CRC64;
 Query Match 94.7%; Score 6923.5; DB 1; Length 1436;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 1361; Conservative 34; Mismatches 41; Indels 1; Gaps 1;
 QY 1 MKDIDIGKEYIIPSPGYRSVREPTSTGTHRDREDSKFRRTRELECODALETAARAGLS 60
 DB 1 MKDIDIGKEYIIPSPGYRSVREPTSTGTHRDREDSKFRRTRELECODALETAARAGLS 60
 QY 61 LDASMSQRLIDEHPKCKYHHGLSALKPRTTCKHQHPVDNAGLFSCFTESWLSLAL 120
 DB 61 LDIVSHSHLQIIDDEHTKGYHHGLSALKPRTTCKHQHPVDNAGLFSCFTESWLSLAL 120
 QY 121 VAHKKGELSMEDVWLSKSHSSDVNCRRLERLMOELNENGPAASLRRVWIFCRTL 180
 DB 121 VAHKKGELSMEDVWLSKSHSSDVNCRRLERLMOELNENGPAASLRRVWIFCRTL 180
 QY 181 LSIYCLMTIQLAGSPGPAFWKHLLEYQATSESNLOYSLVLGLLTLEIVRSMSIALTW 240
 DB 181 LSIYCLMTIQLAGSPGPAFWKHLLEYQATSESNLOYSLVLGLLTLEIVRSMSIALTW 240
 QY 241 ALMYRTGVRLGAILTMAEFKILKLNKIKESLIGELINTCSNDGORMFEAAVAGSLAG 300
 DB 241 ALMYRTGVRLGAILTMAEFKILKLNKIKESLIGELINTCSNDGORMFEAAVAGSLAG 300
 QY 301 PVYALIGMIVNVITLIPPTGLGSAVFLIFPAMMFASRLAYFRRCVAATDRVOKME 360
 DB 301 PVYALIGMIVNVITLIPPTGLGSAVFLIFPAMMFASRLAYFRRCVAATDRVOKME 360
 QY 361 VLYIFIKMYAVKAFVSCQVQKIREERIRILEKAGFQGITGVAPIVYVIAVYFV 420
 DB 361 VLYIFIKMYAVKAFVSCQVQKIREERIRILEKAGFQGITGVAPIVYVIAVYFV 420
 QY 421 HMTLGFDLTAQAFTVYVFNSTFALKVTPESVKSLSASAVALDRFKSLFLMEVHM 480
 DB 421 HMTLGFDLTAQAFTVYVFNSTFALKVTPESVKSLSASAVALDRFKSLFLMEVHM 480
 QY 481 NKRPASHIKIEMKNATLAWSSHSSTONSFKLPKMKKORASRGKREKVRQIQRTEHO 540
 DB 481 NKRPASHIKIEMKNATLAWSSHSSTONSFKLPKMKKORASRGKREKVRQIQRTEHO 540
 QY 541 VLAEOKGHLILDSDEPSPPEEGKHILGHLRLQRTLHIDLEIOGKLVGICGSGSG 600
 DB 541 VLAEOKGHLILDSDEPSPPEEGKHILGHLRLQRTLHIDLEIOGKLVGICGSGSG 600
 QY 601 KTSLSIALIGQMTLLGSLAISGTFAYVAQAAMILNATLNDNILLFGEKDEERYNSVL 660
 DB 601 KTSLSIALIGQMTLLGSLAISGTFAYVAQAAMILNATLNDNILLFGEKDEERYNSVL 660
 QY 661 CCLRPDLAIPNSDLTEIGERGANTSGGQORISLARALSDRSIYILDPALAHVG 720
 DB 661 CCLRPDLAIPNSDLTEIGERGANTSGGQORISLARALSDRSIYILDPALAHVG 720
 QY 721 NHIFNSAIRKHLKSKTVLFTYHOLVYDCEVIFPKKEGCTIRGRGHEELMANGDPA 780
 DB 721 NHIFNSAIRKHLKSKTVLFTYHOLVYDCEVIFPKKEGCTIRGRGHEELMANGDPA 780
 QY 781 FNNLLGEPPEVINSKKESTSGSKSODKGPRTGSIKKKAKVAPGQOLVQLEKQGS 840
 DB 781 FNNLLGEPPEVINSKKESTSGSKSODKGPRTGSIKKKAKVAPGQOLVQLEKQGS 840
 QY 840 FNNLLGEPPEVINSKKESTSGSKSODKGPRTGSIKKKAKVAPGQOLVQLEKQGS 840
 DB 840 FNNLLGEPPEVINSKKESTSGSKSODKGPRTGSIKKKAKVAPGQOLVQLEKQGS 840

```

QY 841 VPMVSVYVYIOAGGPIALVYALFVLMVNGSTAFSTWMLSTYWKSGSGNTTTRGNETS 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 840 VPMVSVYVYIOAGGPIALVYALFVLMVNGSTAFSTWMLSTYWKSGSGNTTTRGNETS 899
QY 901 VSDMKRNPHQVAYSTAYALSMAYMLIKARGVFVKGTLRASRLHDELFRRLILSPM 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 900 VSDMKRNPHQVAYSTAYALSMAYMLIKARGVFVKGTLRASRLHDELFRRLILSPM 959
QY 961 KFEPTPTGRILNFSKDMDEVDVRLPFOAEMFIONVILFECVGMAGVPMFLVAVGP 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 960 KFEPTPTGRILNFSKDMDEVDVRLPFOAEMFIONVILFECVGMAGVPMFLVAVGP 1019
QY 1021 LVILFVSLVHLVSVRLVRLKRLDNIOTSPRLSHITSSIOGLATIHAYNKGOEPLHRIQEL 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1020 LVILFVSLVHLVSVRLVRLKRLDNIOTSPRLSHITSSIOGLATIHAYNKGOEPLHRIQEL 1079
QY 1081 LDDNQAPPELFTCAMRMLAVLADLITALITTTGLMVLVLMHGOILPSAVAGLAISYAVOLT 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1080 LDDNQAPPELFTCAMRMLAVLADLITALITTTGLMVLVLMHGOILPSAVAGLAISYAVOLT 1139
QY 1141 GLFQFTVRLASETEARFTSVYERINHYIKTISLEAPARIKNKAPSPDMPQEGEVTFENAE 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1140 GLFQFTVRLASETEARFTSVYERINHYIKTISLEAPARIKNKAPSPDMPQEGEVTFENAE 1199
QY 1201 RYRENPLVLYLKVVSFTTKPKKIKIGVGTSGKSSGLMALFLVBLVSGGCIKIDGVRISD 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1200 RYRENPLVLYLKVVSFTTKPKKIKIGVGTSGKSSGLMALFLVBLVSGGCIKIDGVRISD 1259
QY 1261 IGLADLSKSLIITIOEPLVFGTVRSNLDPNQYTEDQIMPALEERTHKECIAOLPLKLE 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1260 IGLADLSKSLIITIOEPLVFGTVRSNLDPNQYTEDQIMPALEERTHKECIAOLPLKLE 1319
QY 1321 SEVENNGDNFSGVEROLLICARALLRHCKIILDEPAAAMDETDLLOETIRFADCT 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1320 SEVENNGDNFSGVEROLLICARALLRHCKIILDEPAAAMDETDLLOETIRFADCT 1379
QY 1381 MLIATNHLVTLVGLSDRIMVLAQGVEDPESVLLNSDSFEYMAFAEKVAVK 1437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1380 MLIATNHLVTLVGLSDRIMVLAQGVEDPESVLLNSDSFEYMAFAEKVAVK 1436

RESULT
3
MRP5_MOUSE
ID MRP5_MOUSE STANDARD: PRT: 1436 AA.
AC 09R1X5; 088284;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MUTIDRUG RESISTANCE-ASSOCIATED PROTEIN 5 (ABC TRANSPORTER MOAT-C)
DE (SMRP).
GN ABC5 OR ABC5A OR MRP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20184734; PubMed=10721709;
RA Suzuki T., Sasaki H., Kuh H.J., Agui M., Tatsumi Y., Tanabe S.,
RA Terada M., Saijo N., Nishio K.;
RT "Detailed structural analysis on both human MRP5 and mouse mrp5
RT transcripts.";
RL Gene 242:167-173(2000).
RN [2]
RP SEQUENCE OF 1302-1436 FROM N.A.
RA Suzuki T., Kuh H., Nishio K.;
RC "Molecular cloning of mouse homologue of SMRP/MRP5.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTS AS A MULTISPECIFIC ORGANIC ANION PUMP WHICH CAN
CC TRANSPORT NUCLEOTIDE ANALOGS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY

```

```

CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: AB019003; BAA76609.1; -
DR EMBL: AB012090; BAA32782.1; -
DR MGD: MGI:1351644; Abcc5a.
DR InterPro: IPR001140; -
DR InterPro: IPR001617; -
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT TRANSMEM 179 199
FT TRANSMEM 219 239
FT TRANSMEM 296 316
FT TRANSMEM 317 337
FT TRANSMEM 400 420
FT TRANSMEM 426 446
FT TRANSMEM 608 628
FT TRANSMEM 847 867
FT TRANSMEM 916 936
FT TRANSMEM 996 1016
FT TRANSMEM 1017 1037
FT TRANSMEM 1101 1121
FT TRANSMEM 1126 1146
FT NP_BIND 595 602
FT NP_BIND 1226 1233
FT CARBOHYD 494 494
FT CARBOHYD 636 636
FT CARBOHYD 684 684
FT CARBOHYD 889 889
FT CARBOHYD 896 896
FT CARBOHYD 1043 1043
FT CARBOHYD 1328 1328
FT CARBOHYD 1416 1416
SO SEQUENCE 1436 AA; 161129 MW; 4E32A6E319B8D23C CRC64;

Query Match 93.8%; Score 6855.5; DB 1; Length 1436;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 1354; Conservative 31; Mismatches 51; Indels 1; Gaps 1;

QY 1 MKDDIGKEVYIIIPSPGVSRERSTGTHDRDREKFRRTPRRPECODALFTAAAREGLS 60.
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MKDDIGKEVYIIIPSPGVSRERSTGTHDRDREKFRRTPRRPECODALFTAAAREGLS 60
QY 61 LDASMSQLRIIDEEHPKGYHNGLSALPKPIRTCKKHQHPVDNAGLFSCMTFSWLSLAR 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 LDASMSQLRIIDEEHPKGYHNGLSALPKPIRTCKKHQHPVDNAGLFSCMTFSWLSLAR 120
QY 121 VAHKGGLSMDVWSLSKHESSDVNCRRLERLMOEELNEVGDDAASLRVYVIFCRFLI 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 VAHKGGLSMDVWSLSKHESSDVNCRRLERLMOEELNEVGDDAASLRVYVIFCRFLI 180
QY 121 VVHKKGELMEDVWPLSYKESDVNCRRLERLMOEELNEVGDDAASLRVYVIFCRFLI 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 VVHKKGELMEDVWPLSYKESDVNCRRLERLMOEELNEVGDDAASLRVYVIFCRFLI 180
QY 181 LSVCLMTIOLAGFSGPFWKHLLEYQATESNLOYSLLVGLLREIYRSMALTW 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 LSVCLMTIOLAGFSGPFWKHLLEYQATESNLOYSLLVGLLREIYRSMALTW 240
QY 241 ALNTYGVRLRGAILTMFAFKILIKNIKESIGELINICNSDGOEMFAAAGSLAGG 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 ALNTYGVRLRGAILTMFAFKILIKNIKESIGELINICNSDGOEMFAAAGSLAGG 300
QY 301 PVVAIIIGMIVNYIIIGPGLGSAVFIIIPYAMFASRLTAIFRRKCYAALDERVQKNE 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 PVVAIIIGMIVNYIIIGPGLGSAVFIIIPYAMFASRLTAIFRRKCYAALDERVQKNE 360

```



```

QY 361 VLTYYIKFMVWAKAFSOSVOKIREERRILEKAGYFOSITVGVAVIYVIVASVTSV 420
DB 361 VLTYYIKFMVWAKAFSOSVOKIREERRILEKAGYFOSITVGVAVIYVIVASVTSV 420
QY 421 HMTLGFDTLAAQAFVTVVNSMTPALKVTPSVKSLSEASVAVDRFKSLFMEEVHMIX 480
DB 421 HMTLGFDTLAAQAFVTVVNSMTPALKVTPSVKSLSEASVAVDRFKSLFMEEVHMIX 480
QY 481 NKPAIPHIKEMKNTLWMDSSHSSIONSPLTPPKMKDKRASRGKKEKVRLOLTHQHA 540
DB 481 NKPAIPHIKEMKNTLWMDSSHSSIONSPLTPPKMKDKRASRGKKEKVRLOLTHQHA 540
QY 541 VLAEGKHLLDSDRSPSEEGKHHLGHRLQRTLHSDLEIOEGKLVGICGSVSG 600
DB 541 VLAEGKHLLDSDRSPSEEGKHHLGHRLQRTLHSDLEIOEGKLVGICGSVSG 600
QY 601 KTSLSAILGOMTLLEGSIAISGTFAYVAQAMILNATLNDLILFGEKXEDERSYVLS 660
DB 601 KTSLSAILGOMTLLEGSIAISGTFAYVAQAMILNATLNDLILFGEKXEDERSYVLS 660
QY 661 CCLRPDLATLPSDLTEIGERGANSJGGRORISLARALYSRSTIILDDPSALDAHVG 720
DB 661 CCLRPDLATLPSDLTEIGERGANSJGGRORISLARALYSRSTIILDDPSALDAHVG 720
QY 721 NIIFNSAIRKHLKSTVLFVTHOLOYLVDCEYIFMKKECITERGTHEELMINDYATI 780
DB 721 NIIFNSAIRKHLKSTVLFVTHOLOYLVDCEYIFMKKECITERGTHEELMINDYATI 780
QY 781 FNNLLGEPPEVINSKSKTSQKODKGRGTSIKKEKAVPREGOLVLEEKQGS 840
DB 781 FNNLLGEPPEVINSKSKTSQKODKGRGTSIKKEKAVPREGOLVLEEKQGS 840
QY 841 VWSYVYVYIOAGGELAVLVMALFVLMVNGSTASTMVLSTWIKGSGNTVYTRNENS 900
DB 841 VWSYVYVYIOAGGELAVLVMALFVLMVNGSTASTMVLSTWIKGSGNTVYTRNENS 900
QY 901 VDSMKDNPHMOYASIVASMAVMLILKAIRGVVYVFKGTLRASSLHDELPRILRSM 960
DB 901 VDSMKDNPHMOYASIVASMAVMLILKAIRGVVYVFKGTLRASSLHDELPRILRSM 960
QY 961 KFEEDTPTGRILNRSKMDDEVNLRPQAEFIONVILVFCVGMAGVFWFLVAVP 1020
DB 961 KFEEDTPTGRILNRSKMDDEVNLRPQAEFIONVILVFCVGMAGVFWFLVAVP 1020
QY 1021 LVILSVLHIVSRVILRELKRDNTQSPFLSHITSSIOGLATIAHYNKGOEFLHVOEL 1080
DB 1021 LVILSVLHIVSRVILRELKRDNTQSPFLSHITSSIOGLATIAHYNKGOEFLHVOEL 1080
QY 1081 LDDNAPFLFCAMRWLAVRLDLSIALITTTGLMIVLMHGOIPRAYGLAISVAVOLT 1140
DB 1081 LDDNAPFLFCAMRWLAVRLDLSIALITTTGLMIVLMHGOIPRAYGLAISVAVOLT 1140
QY 1141 GLEFQTVRLASEFAFTSVVERINHYIKTSLSEAPARINKKASPDMPOEGEVTFENAM 1200
DB 1141 GLEFQTVRLASEFAFTSVVERINHYIKTSLSEAPARINKKASPDMPOEGEVTFENAM 1200
QY 1201 RRRENPLVLAKVSVFKIRKEKIGYGRGSGKSSIGMALFRLVELSGCICKIDGIRISD 1260
DB 1201 RRRENPLVLAKVSVFKIRKEKIGYGRGSGKSSIGMALFRLVELSGCICKIDGIRISD 1260
QY 1261 IGLADRSKLSITPOBVFSGTVRSNLDPPNOYTEDQIMDALERTHMEKICIAQLPKLE 1320
DB 1261 IGLADRSKLSITPOBVFSGTVRSNLDPPNOYTEDQIMDALERTHMEKICIAQLPKLE 1320
QY 1321 SEVWENGDNFVGEROLLICARALLHCKILILDEATYAMDTEITDILLIETREAFADCT 1380
DB 1321 SEVWENGDNFVGEROLLICARALLHCKILILDEATYAMDTEITDILLIETREAFADCT 1380
QY 1380 MLIARHLHTVIGSDRIWVLAOGVVEFDPVSLVLSNDSSRFYAMFAAENKVAVVG 1436
DB 1380 MLIARHLHTVIGSDRIWVLAOGVVEFDPVSLVLSNDSSRFYAMFAAENKVAVVG 1436

```

```

RESULT 4
MRP2_RAT STANDARD: PRT: 1541 AA.
AC MRP2_RAT 063120; 063145;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CANALICULAR MULTISPECIFIC ORGANIC ANION TRANSPORTER 1 (MULTIDRUG
DE RESISTANCE-ASSOCIATED PROTEIN 2) (CANALICULAR MULTIDRUG RESISTANCE
DE PROTEIN).
GN ABCC2 OR CMOAT OR MRP2 OR CMRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MISTAR; TISSUE=Liver;
RX MEDLINE=96180672; PubMed=859091;
RA Paulusma C.C., Bosma P.J., Zaman G.J.R., Bakker C.T.M., Otter M.,
RA Schaffer G.L., Scheper R.J., Borst P., Oude Elferink R.P.J.,
RT "Congenital jaundice in rats with a mutation in a multidrug
RT resistance-associated protein gene.";
RL Science 271:1126-1128(1996).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=MISTAR; TISSUE=Liver;
RX MEDLINE=96279006; PubMed=8662992;
RA Buechler M., Koenig J., Brom M., Kartenbeck J., Spring H., Horie T.,
RA Keppler D.;
RT "cDNA cloning of the hepatocyte canalicular isoform of the multidrug
RT resistance protein, CMRP, reveals a novel conjugate export pump
RT deficient in hyperbilirubinemic mutant rats.";
RL J. Biol. Chem. 271:15091-15098(1996).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RA Ito K., Suzuki H., Hirohashi T., Kume K., Shimizu T., Sugiyama Y.;
RT "Expression of the putative ATP-binding cassette region, homologous to
RT that in multidrug resistance associated protein (MRP), is hereditarily
RT defective in Bial hyperbilirubinemic rats (EHRB).";
RL Int. Hepatol. Commun. 292:292-299(1996).
CC -1- FUNCTION: MEDIATES HEPATOBLILIARY EXCRETION OF NUMEROUS ORGANIC
CC ANIONS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE LIVER.
CC -1- DISEASE: DEFECTS IN MRP2 ARE A CAUSE OF HEREDITARY CONJUGATED
CC HYPERBILIRUBINEMIA (EHRB).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL: L49379; AAC62087.1; -
DB EMBL: X96393; CAA65257.1; -
DB EMBL: D86086; BAA13016.1; -
DR HSSP: P13569; INED.
DR InterPro: IPR001140; -
DR pfam: PF00664; ABC_membrane; 2.
DR pfam: PF00005; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT DOMAIN 1 26 EXTRACELLULAR (BY SIMILARITY).
FT DOMAIN 27 47 1 (BY SIMILARITY).
FT DOMAIN 48 67 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 68 88 2 (BY SIMILARITY).

```

| | | | | | |
|----|----------|------|------|-------|-------------------------------------|
| FT | DOMAIN | 89 | 92 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 93 | 113 | | 3 (BY SIMILARITY). |
| FT | DOMAIN | 114 | 125 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 126 | 146 | | 4 (BY SIMILARITY). |
| FT | DOMAIN | 147 | 164 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 165 | 185 | | 5 (BY SIMILARITY). |
| FT | DOMAIN | 186 | 309 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 310 | 330 | | 6 (BY SIMILARITY). |
| FT | DOMAIN | 331 | 356 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 357 | 377 | | 7 (BY SIMILARITY). |
| FT | DOMAIN | 378 | 433 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 434 | 454 | | 8 (BY SIMILARITY). |
| FT | DOMAIN | 455 | 457 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 458 | 478 | | 9 (BY SIMILARITY). |
| FT | DOMAIN | 479 | 540 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 541 | 561 | | 10 (BY SIMILARITY). |
| FT | DOMAIN | 562 | 583 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 584 | 604 | | 11 (BY SIMILARITY). |
| FT | DOMAIN | 605 | 967 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 968 | 988 | | 12 (BY SIMILARITY). |
| FT | DOMAIN | 989 | 1029 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1030 | 1050 | | 13 (BY SIMILARITY). |
| FT | DOMAIN | 1051 | 1093 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 1094 | 1114 | | 14 (BY SIMILARITY). |
| FT | DOMAIN | 1115 | 1115 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1116 | 1136 | | 15 (BY SIMILARITY). |
| FT | DOMAIN | 1137 | 1207 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 1208 | 1228 | | 16 (BY SIMILARITY). |
| FT | DOMAIN | 1229 | 1230 | | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1231 | 1251 | | 17 (BY SIMILARITY). |
| FT | DOMAIN | 1252 | 1541 | | CYTOPLASMIC (BY SIMILARITY). |
| FT | NP_BIND | 667 | 674 | | ATP (POTENTIAL). |
| FT | NP_BIND | 1330 | 1337 | | ATP (POTENTIAL). |
| FT | CARBOHYD | 6 | 6 | | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD | 1007 | 1007 | | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD | 1010 | 1010 | | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBOHYD | 1011 | 1011 | | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CONFLICT | 420 | 420 | | M -> V (IN REF. 3). |
| O6 | SEQUENCE | 1541 | AA: | 17383 | MM; |
| | | | | | MBFDB39 CRC64; |

[illegible]

| | | | |
|----|------|---|-----|
| QY | 467 | FKSLEFMEVMIKKNKPKASPIIK - IEMKNATLADSSHSIIONSPLTPMKKDKRASRG | 525 |
| Db | 609 | IERLYGGDDLLPTSAIRRVSNFDKAVKFESEAFSTD----- | 643 |
| QY | 526 | KKEKVRQLOPTEHQAVLAEQGHLLLDSDERPSPEEREGKHIHGLRLORTLSHIDLEI | 585 |
| Db | 644 | -----PD-----LATIIONVNDI | 657 |
| QY | 586 | QEGKLVGICGSVSGKTSLSIALIGONTLEGSIAISGTPAYAAQAMILNATRONILE | 645 |
| Db | 658 | KPGQLVAVYGVTSKSSLSVAMGEMENVGHITIGGSTAYVPOQSMIONGTIKDNILE | 717 |
| QY | 646 | GKEYDEERYNSVLAISCCRLPDLALIPSDLTEICEBRANISGGORQIRISTARLYSDRSI | 705 |
| Db | 718 | GSEYNEKKYOOVLKACALLPDELITLPGDMAIEKEGKINISGGOKORVSIARAAYODADI | 777 |
| QY | 706 | YILDDPLSALDAHGNHFNENAIKKH - LASKYLVFTYHOLQYIVDCDEVIFMKEGCITE | 763 |
| Db | 778 | YILDDPLSVAVHAHGKHLFNKVVPNGLLAKTRIEFVTHGHELPQYDEIVLVGKGTILE | 837 |
| QY | 764 | RGTHEELMNLNGDPAATTFNNLLGEPFVEI----- | 794 |
| Db | 838 | KGSYDLDLKKGVYARANKMKTMTKISGEBGEATVYNDSDEADDDDGILPYHEELPEDAASL | 897 |
| QY | 795 | -----NSKREY - SGSOKKQODGP-----KTGSJIKKRAYKPEBGOLVOLEKGOGS | 840 |
| Db | 898 | AMRRENSLRFTLSSSSSSSSRRGSLKNSLKIKNVNLKKEKEVEBQKLIKKEPEYETGK | 957 |
| QY | 841 | VPWYSYGYITOAAGGPLAFVYALFMNLNNGSTAFSTWMLSTYMKOGSGTFTYTRGNETS | 900 |
| Db | 958 | VKFSYLYLQAVGM - WSILPILILEFGINNVAFTGSMILMSAMTSD - SDMI-----NGTN | 101 |
| QY | 901 | VSDSKDNPRHMOYVASTYALSAVMLILKIRGVVFEVFGKL-----RASSRLHDELFR | 953 |
| Db | 1011 | NSSSHRD-----MRTGFGALGAQIGCICLILSTLMSIYACRNASKALHGOULLT | 105 |
| QY | 954 | RILRSPKPFPTTPTGRIILNFSKDMDEVDVRLPFOAMEFIQWYILFEFCVG-----MLAG | 100 |
| Db | 1059 | NILRAPMFEPTPTPTGRIVNFSGDISIYDOLLP-----QILRSMWMOFCIATIVIMIC | 111 |
| QY | 1010 | VEPWFVAVGVPLFESVLAHVSRYLIRELKRDLNITQOSFPLSHITSSIOGLATHTAYNK | 106 |
| Db | 1115 | ATPFAAIIIEPLSLIYSVOVFYVATSRQOLRRDLSYKSPYSHFSEYVGLPIRAFHEH | 117 |
| QY | 1070 | GOEFHRYOELLDNQAPFLFTFCAMBRILVRLDISALITITTGMIYVLMHQIDPRATA | 112 |
| Db | 1175 | QORFLANNEKQIDINQCKVSWITSNMLAIRLEVLNVLFGSALVLYRYKTLTGIDVY | 123 |
| QY | 1130 | GLAISYAVOULTGFQFVYRLASETEARFTSYERLNHKTLSLEAPARIKNKAPSDWQ | 118 |
| Db | 1235 | GFVLSNALNITQTLNMLVLRSEAFETNIVAVERISEYI - NVENBAP - WYTDKRPADWDR | 129 |
| QY | 1190 | EGBEYTFENAEKRYEENCLPYLUKKVYSPFIKPEKIGIVGRTGSGKSSIGMALPFLVELSGG | 124 |
| Db | 1293 | HGEIOFNMYQVRYREPLDLYLKGITCINIKSGEKKYGVGRTGACKSSITLNCFLRILESAGG | 135 |
| QY | 1250 | CIKIDGVASIDIGLADIRSKSIIIRQBPVLFSGYVRNLDLPFQNYTEDQYIWDALFETHNK | 130 |
| Db | 1353 | QIITIDGIDVAGIHLHDERLERTIIPOLILFSSGLLRNMLDPFNKYISDSEEWRALELAHFR | 141 |
| QY | 1310 | ECIAOLPFLTESEWENGDNFSVEGRDOLCLARLYLHNCKLILDEXTAMPDETLLIO | 136 |
| Db | 1413 | SFVSGLOGLLSEYTBEGDNLSTIGORLCLGRVLRKKSIIYVIDERTAAVDELFTSLIO | 147 |
| QY | 1370 | ETIREAPADCTMLTIAIHLRTLVGSDRIMVLAQGVYVEFTPSVLLSNDSSRFYAMF - A | 142 |
| Db | 1473 | TTIRKFSOQCTVITIRARLRTIMDSKIXIWLNDGKIIVEYSGPELLSNRGS - FYLMAXEA | 153 |
| QY | 1428 | AAEN 1431 | |
| Db | 1532 | GIEN 1535 | |

```

RESULT 5
ID MRPL_HUMAN STANDARD: PRT: 1531 AA.
AC P33527.014819; F78419;
DT 01-FEB-1994 (Rel. 28; Created)
DR 01-NOV-1997 (Rel. 35; Last sequence update)
DE 01-OCT-2000 (Rel. 40; Last annotation update)
DE MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN 1.
GN ABC1 OR MRP OR MRPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9308080; PubMed=1360704;
RA Cole S.P.C., Bhadwaj G., Gerlach J.H., Mackie J.E., Grant C.E.,
RT Almqvist K.C., Stewart A.J., Kurz E.U., Duncan A.M.V., Deeley R.G.;
RT "Overexpression of a transporter gene in a multidrug-resistant human
RT lung cancer cell line."
RL Science 258:1650-1654 (1992).
RN [2]
RP REVISIONS.
RX MEDLINE=93262415; PubMed=8098549;
RA Cole S.P.C., Deeley R.G.;
RT "Multidrug resistance-associated protein: sequence correction."
RL Science 260:879-879 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96251691; PubMed=8649356;
RA Stride B.D., Valdimarsson G., Gerlach J.H., Wilson G.M.,
RT Cole S.P.C., Deeley R.G.;
RT "Structure and expression of the messenger RNA encoding the murine
RT multidrug resistance protein, an ATP-binding cassette transporter."
RL Mol. Pharmacol. 49:962-971 (1996).
RN [4]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=98008927; PubMed=9344662;
RA Grant C.E., Kurz E.U., Cole S.P.C., Deeley R.G.;
RT "Analysis of the intron-exon organization of the human multidrug-
RT resistance protein gene (MRP) and alternative splicing of its mRNA."
RL Genomics 45:368-378 (1997).
RN [5]
RP SEQUENCE OF 1131-1531 FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.U., Kim U.-J., Shedd V.P., Kalush F., Brandon R.,
RT Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RT Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q."
RL Genomics 60:295-308 (1999).
RN [6]
RP TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=97442425; PubMed=9295302;
RA Hifner D.R., Almqvist K.C., Leslie E.M., Gerlach J.H., Grant C.E.,
RT Deeley R.G., Cole S.P.C.;
RT "Membrane topology of the multidrug resistance protein (MRP). A study
RT of glycosylation-site mutants reveals an extracytosolic NH2
RT terminus."
RL J. Biol. Chem. 272:23623-23630 (1997).
RN [7]
RP TOPOLOGY.
RX MEDLINE=97476249; PubMed=9334225;
RA Kast C., Gros P.;
RT "Topology mapping of the amino-terminal half of multidrug resistance-
RT associated protein by epitope insertion and immunofluorescence."
RL J. Biol. Chem. 272:26479-26487 (1997).
RN [8]
RP TOPOLOGY.
RX MEDLINE=98153110; PubMed=9485377;
RA Kast C., Gros P.;
RT "Epitope insertion favors a six transmembrane domain model for the

```

```

RT carboxy-terminal portion of the multidrug resistance-associated
RT protein."
RL Biochemistry 37:2305-2313 (1998).
CC -!- FUNCTION: MAY PARTICIPATE DIRECTLY IN THE ACTIVE TRANSPORT OF
CC DRUGS INTO SUBCELLULAR ORGANELLES OR INFLUENCE DRUG DISTRIBUTION
CC INDIRECTLY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: LUNG, TESTIS, AND PERIPHERAL BLOOD
CC MONONUCLEAR CELLS.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05628; AAB46616.1; -
DR EMBL; AF022853; AAB83983.1; -
DR EMBL; AF022824; AAB83983.1; JOINED.
DR EMBL; AF022825; AAB83983.1; JOINED.
DR EMBL; AF022826; AAB83983.1; JOINED.
DR EMBL; AF022827; AAB83983.1; JOINED.
DR EMBL; AF022828; AAB83983.1; JOINED.
DR EMBL; AF022829; AAB83983.1; JOINED.
DR EMBL; AF022830; AAB83983.1; JOINED.
DR EMBL; AF022831; AAB83983.1; JOINED.
DR EMBL; AF022832; AAB83983.1; JOINED.
DR EMBL; AF022833; AAB83983.1; JOINED.
DR EMBL; AF022834; AAB83983.1; JOINED.
DR EMBL; AF022835; AAB83983.1; JOINED.
DR EMBL; AF022836; AAB83983.1; JOINED.
DR EMBL; AF022837; AAB83983.1; JOINED.
DR EMBL; AF022838; AAB83983.1; JOINED.
DR EMBL; AF022839; AAB83983.1; JOINED.
DR EMBL; AF022840; AAB83983.1; JOINED.
DR EMBL; AF022841; AAB83983.1; JOINED.
DR EMBL; AF022842; AAB83983.1; JOINED.
DR EMBL; AF022843; AAB83983.1; JOINED.
DR EMBL; AF022844; AAB83983.1; JOINED.
DR EMBL; AF022845; AAB83983.1; JOINED.
DR EMBL; AF022846; AAB83983.1; JOINED.
DR EMBL; AF022847; AAB83983.1; JOINED.
DR EMBL; AF022848; AAB83983.1; JOINED.
DR EMBL; AF022849; AAB83983.1; JOINED.
DR EMBL; AF022850; AAB83983.1; JOINED.
DR EMBL; AF022851; AAB83983.1; JOINED.
DR EMBL; U91318; AAC15784.1; -
DR PIR; A44231; DVHUAR.
DR HSSP; P13569; INBD.
DR MIM; 158343; -
DR InterPro; IPR001140; -
DR InterPro; IPR001617; -
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Alternative splicing.
FT DOMAIN 1 33 EXTRACELLULAR.
FT TRANSMEM 34 54 1.
FT DOMAIN 55 74 CYTOPLASMIC.
FT TRANSMEM 75 95 2.
FT DOMAIN 96 100 EXTRACELLULAR.
FT TRANSMEM 101 121 3.
FT DOMAIN 122 133 CYTOPLASMIC.
FT TRANSMEM 134 154 4.
FT DOMAIN 155 172 EXTRACELLULAR.
FT TRANSMEM 173 193 5.

```

```
FT DOMAIN 194 316 CYTOPLASMIC.
FT TRANSMEM 317 337
FT DOMAIN 338 363 EXTRACELLULAR.
FT TRANSMEM 364 384
FT DOMAIN 385 440 CYTOPLASMIC.
FT TRANSMEM 441 461
FT DOMAIN 462 464 EXTRACELLULAR.
FT TRANSMEM 465 485
FT DOMAIN 486 547 CYTOPLASMIC.
FT TRANSMEM 548 568
FT DOMAIN 569 590 EXTRACELLULAR.
FT TRANSMEM 591 611
FT DOMAIN 612 967 CYTOPLASMIC.
FT TRANSMEM 968 988
FT DOMAIN 989 1025 EXTRACELLULAR.
FT TRANSMEM 1026 1046
FT DOMAIN 1047 1089 CYTOPLASMIC.
FT TRANSMEM 1090 1110
FT DOMAIN 1111 1113 EXTRACELLULAR.
FT TRANSMEM 1112 1132
FT DOMAIN 1133 1203 CYTOPLASMIC.
FT TRANSMEM 1204 1224
FT DOMAIN 1225 1226 EXTRACELLULAR.
FT TRANSMEM 1227 1247
FT DOMAIN 1248 1531
FT NP_BIND 678 685 ATP (POTENTIAL).
FT NP_BIND 1327 1334 ATP (POTENTIAL).
FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .).
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .).
FT CARBOHYD 1006 1006 N-LINKED (GLCNAC. . .).
FT VARSPPLIC 706 764 MISSING (IN ISOFORM WITHOUT EXON 17).
FT VARSPPLIC 765 820 MISSING (IN ISOFORM WITHOUT EXON 18).
FT VARSPPLIC 1431 1495 MISSING (IN ISOFORM WITHOUT EXON 30).
SQ SEQUENCE 1531 AA; 171560 MW; 5CCB92FEB7B096B52 CRC64;
```

Query Match 28.5%; Score 2082; DB 1; Length 1531;

Best Local Similarity 34.5%; Pred. No. 1.4e-127; Matches 488; Conservative 262; Mismatches 479; Indels 184; Gaps 21;

```
100 PVDNAGLESCMTFMSLSLARVAHKKGELSMEDWLSKSHSSDVNCRBLRLMOEELNE 159
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
209 PESSASPLSRITFTFWITLRI-VRGYRQPLBESDLINKEDTSEGVVYVLYKMKKECAK 267
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
160 V-----GPDAA-----SLRRVWIFCTRLIL 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
268 TRQPVKKVYSSKDPACQKSSKYDANEVEALIVKSPQKEMNPISLFKYLKTEGPFILM 327
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
182 SITCLMTIQLAGFSGPAMVKKHLEIYQATSNIOYSLLVGLLTITIVSMGLATLWA 241
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
328 SEFFKAIHDMFMFSGPO-ILKLILKRVNDTKAPWQGYFTVLLFVTAQLDTLVLYHOYFH 386
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
242 LNYRGTVALRGALITLMAKKTILKLNKERS--LGEIINISNDGQRMFEAAVAGSLIAG 299
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
387 ICFPSGRKIKTAIVGAVYKRALVITNSARKSSYGEIYNLMSVDAQRMDLATYINMIMS 446
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
300 GPVVAIIIGMIYVNIIGTGFGLSVAFLIFYPAMMFASRLTAFFRRKCAVATDERVQKMN 359
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
447 APLOVLIALLYLMINTLGPVLAGVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 506
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
360 EVLYTYIKFYKAYVAKFASQVOKIREERKILEKAGFQSIYGVAVIYVAVIYVAVIFS 419
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
507 EILNGIKVLYKLYAMELAKFDVLAIRQBELVLYKLSAYLSAVGFTTWCTPFLVALCTFA 566
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
420 VHMILGPD--LTAQAQFVTVVNSMTPALKVTPSVKSLSEASVAVDREFSLMEEVH 477
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
567 VYVITDERNNIIDAQTAFAVSLFLNLRPLNLRPLNLRPLNLRPLNLRPLNLRPLN 626
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
478 --MIKNRPASP---HIKEMNATLAMDSSSHSISQNSPKLPPKMKDKRASRGGKREKVRQ 532
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
627 PDSIERBRVVKVGGGNSITVNAFAFTW----- 653
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
533 LQRTBHQAVLAEQGHLLDSDERPSPPEEKGKIHHLQRTLQRTLQRTLQRTLQRTLQ 592
```

```
Db 654 -----ARSDP-----TINQITSIPEGALVA 675
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
593 ICGVSGKTSLLSALLGQMTLLGSLAISGTFAYVAQOAILNATLNDILFKEXDEY 652
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
676 VVGQVGGCKSSLLSALLAEMDKVEGHVAKISGVAVVPOQAMIONDSIRENITLFCQLEEP 735
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
653 RYNSVNSCCLRPDILATIPSSDLTEIGRGANLSGGQQRQISLARALYSDSIYIILDDPL 712
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
736 YKRSYIQAICALLPDLLEILPDSGRTEIGKGVNLSGGQKQKRSARAVYSNADYILFPDPL 795
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
713 SALDAHVNHIENSAT--RKHLKSKTVLFVTHOLOYVDCDEVIFMEKGCITERTGTHEEL 770
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
796 SAVDAHVCKHIFENVYIGKGLKKNKTRILTYHMSYLPQVQVVIYMSGKISEMGQVEL 855
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
771 MNLNGDYATIF-----NLL-----GETPPYEIN 795
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
856 LARQAFAPFLRTYASTRQODAEENGVTGVSGPKKEAKQKMGKLVDSAGKOLQRLS 915
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
796 SKKETSGQKKSQDKGPKTSGIKKEKAVKPREGOLVLEEGGQSVPMYGVYIQAAG 855
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
916 SSSYSQDISRHHN--STAELQKAERKEETWKLMEADKQGTQVKLSYWDYWKALIGL 972
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
856 PLAFVIALFPLNNGVSPAFSTMWLSYIKQSGNTVTPRGNETSVSDMKDNPHMOYVA 915
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
973 FISFLSIF-LFMCHNVSLAANYMLSLMTDDPIVNGT--QEHKVRSLSYGALGISGI 1028
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
916 STYALSAVMMLILKAIKNGVFEVKGTLRASSRLHDELFRILRSPKKEFTPTPTGRIILRF 975
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1029 AVFGYSAVSI-----GGIILASRCHLDLISILRSPSPFEPFSGNLVRF 1076
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
976 SKMDDEVYRPLPQAKEMFIQVAVILYFPCVGMAGVFPFPLVAVGPIVLEFVLIHYSVL 1035
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1077 SKEDTVDSMIPEVKIKMKGSLFVNYGACIVILATPAAIILPILGLITY--FFVQREY 1133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1036 I---RELKRLNITQSPPLSHITSSIOGLATTHAYNKGQEFLLHRYOELDDNOAPFLFT 1092
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1134 VASSQQLKRLSSVSNPSYSHFNETILGVSYIRAFEBBERITHQSDLKVDENQKAYYSI 1193
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1093 CAMRLAVRLDLISALITTTGLMIVLMHGOIPRAYAGLAISYAVOLTLGFQFVRLASE 1152
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1194 VANRLAVRLBEOVCNIVLFAALFAVIRSHSLAGVLGLSVSYLQVTTYVNMVLRMSE 1253
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1153 TEARTSVIRINHYIKTSLAPAIKKAAPSPPMPQGEVTFEAKERYRENPLVLYK 1212
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1254 METNIVAEERLKEYSET-EKAPWQIOETAPSSWPQVGRVEFRNYKCLRYREDLDFVLRH 1312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1213 VSFITKPKKXIGVGRSGSSLSGMALFRLVELSGGCIKIDGVIRISDIGLADLRSLKSI 1272
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1313 INVITNGEKEKGIYGRTAGSSILTLGLFRINESAGEIITIDGIMAKIGLHDLRFKTI 1372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1273 IPOEVLFGSTVRSNLDPFNOYTEDOIMDALERTHMKECIAQLPLKLESEVMENGDFSV 1332
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1373 IPQDVLFGSGLRMNLDPFQSYSDDEVMTSLELAHLKDFVSALPDKLDHECAEGENISV 1432
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1333 GEROLCTARALLRCKILLIDEPATAMDTEDDLIOETINEAPFDCIMLTIAHRLHYL 1392
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1433 GQROLYVCLARALLRCKTKLILDEAPAAVDELDDLIQSTIRTFQFEDCTVLTIAHRLTIM 1492
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1393 GSDRIMVLAOGOVVEFDPVSVLSDSSRYFAM 1425
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1493 DTYRIVLDKGEIDETYGAPSDLL-QQRLGFLYSM 1524
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 6

MRP2_HUMAN STANDARD; PRT: 1545 AA.
ID MRP2_HUMAN Q92887; Q92863; Q92798; Q14022; Q92500;
AC Q92887; Q92863; Q92798; Q14022; Q92500;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CANALICULAR MULTISPECIFIC ORGANIC ANION TRANSPORTER 1 (MULTIDRUG
RESISTANCE-ASSOCIATED PROTEIN 2) (CANALICULAR MULTIDRUG RESISTANCE


```

QY 194 FSGP---AFWVHLEXT-----GATESNLOYSLLVIGLL-----TE 229
Db 302 FTQPOLRLIKETVYDSEKODHSSLOGFENHPKLTIVRGFLIAPFAMELVGTOTS 361
QY 230 IVRMSALATLALVYRGVRLGALITFMAFKILKLNKE--KSLGELINCSNDGORM 287
Db 362 VLHQYFLNV-----FNGMYIKSALTALITYOKSLVLSNEAGSLSTGDIIVLMSTVDYKL 416
QY 288 FEAAVAGSLAGVVALIYVNIILGPTGIGSAVFILFYPAMPASRLTAYFRKRC 347
Db 417 QDLQWMLNLSGPFOLITCLYSYLKLGNSMWGVILVIMPLNPSLMIQOKLKQSKQ 476
QY 348 VAADERQKNNVEITYIKFKMTAMVAFSOSQKIR--EERRRLKAGFQOSTTVEVA 406
Db 477 MKYDERTRVISELLNNKSKLXAMEKPREKLEEVKNNKELNLTJLGCMATTSFOF 536
QY 407 PIYVIVASVWFESVHM--TIGFDLTAAGAFVTVFVNSMTFLKVPFVSXLSSEAVYD 465
Db 537 NIVPELVSCFFAVFVYTEDRALTTDLVFPALTLEFNLISFPLMITPMVINSFTEASVIG 596
QY 466 RFSKSLFMEVYHMKRNPASPHIKEMKATLANDSSHSSIONSFKLTPKMKKKRASRG 525
Db 597 RLFFFTNEEL-----QPDVQRLPKV--KNIGVAINIG 629
QY 526 KKEVROLQRTREHQAVALAEOKGHLILDSDEKSPREECKHILGHLKQLSHIDLEI 585
Db 630 DDAFELMORKPEYKV-----ALKNNINQA 653
QY 586 QEGKLVIGCVSGSKTSLIATIGOMTLLEGSAISGTPAVYAAQOATLNTLNDITLF 645
Db 654 KKNLITCIVGKSGKTALISCMGLDLEFRVKGFAVHGSVAVSVPIMMGTVENILF 713
QY 646 GKEYDEERYNSVLNSCCLRPDLALIPSSDLTEIGERANLSGGQORISIALRALSDBSI 705
Db 714 GHRDAEYEKTIKACALITDILAIIMDGDKTLVEKGISLSCGQARISLARAVYARADT 773
QY 706 YILDDPLSALAHQVNHIFNSAIRKH--LKSQTVLFVTHQOYLVDCEVIFMKGCCTE 763
Db 774 YLLDDPLAAVEHVARHLEHYLGPNGLHRTKVLATNKSALSIADSIALLDGEITQ 833
QY 764 RGTHEEL--MNLNGDYATIFENLLIGETPPVEINSKETSQSGOKSDQPKGTGSKKEKA 822
Db 834 QGTDEITRKDSDPLMKLLNNY-----GKKNNG--KSENGDSESSSVRESS 878
QY 823 VKPEGQULVOL-----EKGQGSVPMS 844
Db 879 I-PVSGELEQOKLUNDLDFGNSDAISLRASDATLGSIDFGDENIARKEHQKVAWN 937
QY 845 VYGVYIOAAGPIAFVLYMALEFNLNVGSAFSTWMLSTWIKOGENNTVTYKGNESVSDS 904
Db 938 IYLEYAKAC--NPKSVCFILFIVISMFLSVGNVWLKHM-----SEVNSR 981
QY 905 MKDNPMMOYASIV---ALSMAYMLILKAIGVVPVKGTLRASRLHDELFRRIIRSPMK 961
Db 982 YGSNNNARIYALIVNADIGSALFLIOTI--VLMVECTIASKSLHNLNMTNSVYARAPT 1039
QY 962 FPDTPTRIGIRIRESKMDVEDVRLPFOAEMFIQVNIIVFCVGMIAGVFWELVAVGPL 1021
Db 1040 FEETTPIRIIRIRESNDIYKVDALLGRFSQFVNAVKVTFITIVICATWQFIILPL 1099
QY 1022 VILFVLAHIVSRVILREIKRLDNTOSPFILSHITSSIOGLATIHAYNKGQELHYOELL 1081
Db 1100 SVFYIYQOYLRKTSRKRDSITRSPYSHFOETLGLAVRGVSOOKRSHINOCRI 1159
QY 1082 DDNOAPFLFTCAMFLAVRLDLISALITTTGLMIV--LMHGQPPPAYAGIAISYAVOL 1139
Db 1160 DNNNSAFYPSINANWMLVRLLEIGSIIILGAATLSVRLKGTITLAGVNGSLSTYALQI 1219
QY 1140 TGLFOFVRLASEFARTSVYRINHYYKTLSLEAPARIKNAKSPDMPQSEVTFENAE 1199
Db 1220 TOTLNMIVRMVTEVETNIVSVRIKEY--ADLKSEAPLIVEGHRPKEWPSQDIFKNYNS 1278
QY 1200 MRYRENLPVLVKKVSFTIKPEKIGIVRGTSGSKSLGMAFLRLVLESGGCKIKIDGVRIS 1259

```

```

Db 1279 TRYREEDLVKHNINHKPKENKGVIGRTGAGKSSLLALFPMIEASEGNIVIDNIAIN 1338
QY 1260 DIGLADLRKSLIIPQEVLFSGVYRSLMDPPNOYTEDQINDALERTHMKACIAQLPK- 1318
Db 1339 EIGLYDLRHKLSIIPQDSQVFEGRVRENIDPINOYTDIAIRALELSHKLEHVLSMSMDG 1398
QY 1319 LSEFVEMGDNFVSGBERDLICARALLRHCKLITIDETAAMDQETDILLIOTETREAPAD 1378
Db 1399 LDAQLTGEGGMLSVGQROLICLARMLVPSKILVLEDAVTAVDVETKVOVQETITAKFD 1458
QY 1379 CMTLTAIRLHTVLGSDRIWYLAQGOVEFDPVSVLSDNSSRFYAM 1425
Db 1459 RTIILTAIRLMTIMDSRITIVLDNKAFAFDSPOGLLSDNKSIFYSL 1505

RESULT 10
MRP3_RAT STANDARD: PRT: 1522 AA.
ID MRP3_RAT 088563;
AC 088563; 088270;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CANALICULAR MULTISPECIFIC ORGANIC ANION TRANSPORTER 2 (MULTIDRUG
DE RESISTANCE-ASSOCIATED PROTEIN 3) (MRP-LIKE PROTEIN-2) (MLP-2).
GN ABCC3 OR CMOAT2 OR MRP3 OR MLP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99292429; PubMed=10362653;
RA Ortiz D.F., Li S., Iyer R., Zhang X., Novikoff P., Arias I.M.;
RT "MRP3, a new ATP-binding cassette protein localized to the canalicular
RT domain of the hepatocyte."
RL Am. J. Physiol. 276:G1493-G1500(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Colon;
RX MEDLINE=98279126; PubMed=9614210;
RA Hichashi T., Suzuki H., Ito K., Ogawa K., Kume K., Shimizu T.,
RA Sugiyama Y.;
RT "Hepatic expression of multidrug resistance-associated protein-1 like
RT proteins maintained in eisel hyperbilirubinemic rats."
RL Mol. Pharmacol. 53:1068-1075(1998).
CC -1- FUNCTION: MAY ACT AS AN INDUCIBLE TRANSPORTER IN THE BILIARY AND
CC INTERSTINAL EXCRETION OF ORGANIC ANIONS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: LUNG AND INTESTINE, LOW IN LIVER. HIGHER IN
CC LIVER OF EISEL HYPERBILIRUBINEMIC RATS AND TR(-) MUTANT RATS.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL, AF072816; AAC25416.1; -
CC EMBL, AB010467; BAA28955.1; -
CC HSSP: P13569; INED.
CC InterPro: IPR001140; -
CC InterPro: IPR001617; -
CC Pfam: PF00664; ABC_Membrane; 2.
CC Pfam: PF00005; ABC_Tran; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT DOMAIN 1 32 EXTRACELLULAR (BY SIMILARITY).

```

| | | | | |
|----|----------|----------|-----------------------------|---|
| FT | TRANSMEM | 33 | 53 | 1 (BY SIMILARITY). |
| FT | DOMAIN | 54 | 73 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 74 | 94 | 2 (BY SIMILARITY). |
| FT | DOMAIN | 95 | 99 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 100 | 120 | 3 (BY SIMILARITY). |
| FT | DOMAIN | 121 | 132 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 133 | 153 | 4 (BY SIMILARITY). |
| FT | DOMAIN | 154 | 171 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 172 | 192 | 5 (BY SIMILARITY). |
| FT | DOMAIN | 193 | 301 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 302 | 322 | 6 (BY SIMILARITY). |
| FT | DOMAIN | 323 | 347 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 348 | 368 | 7 (BY SIMILARITY). |
| FT | DOMAIN | 369 | 424 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 425 | 445 | 8 (BY SIMILARITY). |
| FT | DOMAIN | 446 | 448 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 449 | 469 | 9 (BY SIMILARITY). |
| FT | DOMAIN | 470 | 531 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 532 | 552 | 10 (BY SIMILARITY). |
| FT | DOMAIN | 553 | 574 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 575 | 595 | 11 (BY SIMILARITY). |
| FT | DOMAIN | 596 | 958 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 959 | 979 | 12 (BY SIMILARITY). |
| FT | DOMAIN | 980 | 1016 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1017 | 1037 | 13 (BY SIMILARITY). |
| FT | DOMAIN | 1038 | 1080 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 1081 | 1101 | 14 (BY SIMILARITY). |
| FT | DOMAIN | 1102 | 1123 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1103 | 1123 | 15 (BY SIMILARITY). |
| FT | DOMAIN | 1124 | 1194 | CYTOPLASMIC (BY SIMILARITY). |
| FT | TRANSMEM | 1195 | 1215 | 16 (BY SIMILARITY). |
| FT | DOMAIN | 1216 | 1217 | EXTRACELLULAR (BY SIMILARITY). |
| FT | TRANSMEM | 1218 | 1238 | 17 (BY SIMILARITY). |
| FT | DOMAIN | 1239 | 1522 | CYTOPLASMIC (BY SIMILARITY). |
| FT | NP_BIND | 659 | 666 | ATP (POTENTIAL). |
| FT | NP_BIND | 1318 | 1325 | ATP (POTENTIAL). |
| FT | CARBOHYD | 18 | 18 | N-LINKED (GLYCANC. . .) (POTENTIAL). |
| FT | CARBOHYD | 1001 | 1001 | N-LINKED (GLYCANC. . .) (POTENTIAL). |
| FT | CARBOHYD | 1002 | 1002 | N-LINKED (GLYCANC. . .) (POTENTIAL). |
| FT | CONFLICT | 323 | 344 | SSSTHSCASSSGFLRPHGY -> LSFINPOLLSILIRNF |
| FT | CONFLICT | | | ISDPAAP (IN REF. 2). |
| FT | CONFLICT | 645 | 645 | I -> L (IN REF. 2). |
| FT | CONFLICT | 1075 | 1075 | H -> D (IN REF. 2). |
| Q0 | SEQUENCE | 1522 AA: | 168977 MW: 74035180C4C64297 | CRC64; |

| | | | |
|----|---|----------------------------|--------------|
| | Query Match | 26.8%; Score 1960; DB 1; | Length 1522; |
| | Best Local Similarity | 32.9%; Pred. No. 1.2e-119; | |
| | Matches 473; Conservative 261; Mismatches 463; Indels 240; Gaps 33; | | |
| QY | 100 PVDNAGLFCSCMTFSWLLSSLARAHAKKGELSMEDVSLSKSHESDYNCRFLERLMQEELNE | 159 | |
| | : : : : : : : : : : : : : : : : : : : | | |
| Db | 208 PEASAGFGRSRLSFWWFPTKLALIGYR - PLBESDLMSLESDCSHKRYQVQLLEAMKQQRNQ | 266 | |
| QY | 160 V-GPDAS-----LRRVNIFFCTRLILSLIVCIMITQ---- | 190 | |
| | : : | | |
| Db | 267 ASGPOTALPEPKAGEDVELLKAREPTEKKPSPLRALVFET--TSSLIMGACEFKLIQDLSLP | 324 | |
| QY | 191 -----LAGFSRP--APAVKHLLLETQATFESTL-----QYSLLLVGLLTLEY | 231 | |
| | : : : : : : : : : : : : : : : : : : : | | |
| Db | 325 SSTHSCASSSGGLEPRHPRGVWGFFLLAGLM-FVSSTMOPLLHOHYHCIFVAL----- | 377 | |
| QY | 232 RWSMALTLWALNYFRGVRLRGAILTMARFKILTKLNI--KEKSIGELINICNDGORMPE | 289 | |
| | : : : : : : : : : : : : : : : : : : : | | |
| Db | 378 -----RIKRALIIQVIIRKALITITSNKRETYVEEMVLMSVDMDQRFMD | 420 | |
| QY | 290 AAANGSLLAGGEVVAALLGMIVNVILLGFPTGLGSAAVFLIEYPAMMFASRLTAYPEFRKCYA | 349 | |
| | : : : : : : : : : : : : : : : : : : : : | | |
| Db | 421 VSPFNILMASPLQYITIAIFYELMQLLGPSALAGAVIYLLIDPLNGAVNSMKMTQYQYQXK | 480 | |
| QY | 350 ATDENAVQAMNEVLITYIKRTIKTAWKKAISQSQAKTREEERRILLEAGRYQOSTITVGAPLY | 409 | |
| | : : : : : : : : : : : : : : : : : : : | | |
| | 481 EKDSIKIKMSIELMIQVKLYKLVIAAMPPTLEOVGGIRGGELQLLRGAYIQDAISTFTPWCT | 540 | |

| | | | |
|----|------|--|------|
| QY | 410 | VVLSVSVTFSSHMLTGD--LTAQAQFVTVTVNSMTFLAKTTPPSVYSLSSASVANDR | 467 |
| Db | 541 | PFNWTLITLGYVCVDKNNVNLDAEKAFVSLTFENLKLPLNLPLPOLISQMTQVSLKRI | 600 |
| QY | 468 | KSLFLMEEV--HMKNKPSAPHIKIEKNAKTLAMDSHSSSIONSPRLPKMKDKRASR | 525 |
| Db | 601 | QDFLMDLDPQVCKRTTISPGRATIIHNGT\$W-----SKD----- | 637 |
| QY | 526 | KKEVRQLOFTEHOAVLAEOKGHLHDSDERSPEEBEKGKHLGLRLQRTLSIDEI | 585 |
| Db | 638 | -----LPLTHSINIQI | 649 |
| QY | 586 | QEGKIVGICSVSGKTSLSIALIGQMTLLEGSIAISGFPAVQAQWALNTLPDNLF | 645 |
| Db | 650 | PKGALVAVAGVPGCGKSLVSALLGEMKELEGAVSKGVAAVPOQAMIONCTLOENYLF | 709 |
| QY | 646 | GKEVDERKYNVLSNCCLRDLALIPSQDTEIGEAGANLGGORQRLSLAALQSDSI | 705 |
| Db | 710 | GQPMNPKYQOALETCALLADLDVLPBGDQTEIGERKIMLSGORQRLSLAAVYSDANI | 769 |
| QY | 706 | YTLDDPSALDAHAGNIIFNSAIRKH--LKSQVLELVTHQLOLYDCCVIMPKAGCITE | 763 |
| Db | 770 | FLDDPELSADSHAKIIPQVIGPEVLAGKRVLYTHGISLPLQTDPIIYLAGQITE | 829 |
| QY | 764 | RGTHEELMNLNGDYATIFNN-----LLG\$T----- | 791 |
| Db | 830 | MGRVSEILLQHDG\$PANFLRVNAPDENQANEGVLQHANEBVLLLEDTLSTHDTLDTP\$A | 889 |
| QY | 792 | V-EIN\$KKEITSGSK\$Q\$D\$GPK-----T\$GIKE--KAPKREGQVQLEEKQ\$Q\$SY | 841 |
| Db | 890 | IYEVAKQFMRM\$SLSEBGGQ\$NRPVLKRYT\$SL\$E\$K\$E\$V\$P\$ATQ\$K\$E\$G\$AL\$K\$E\$IA\$E\$TV\$N | 949 |
| QY | 842 | PMSVYGVYIOAAG\$PLAEVLVIML\$MNV\$STP\$F--\$TM\$LM\$SY\$M\$K\$Q\$G\$TV\$TR\$G-- | 896 |
| Db | 950 | KLSYIMDYAK\$V\$---LCTTFLPICLIYAQ\$NVAIGANVLSAMT-----NDVEHQQ | 1000 |
| QY | 897 | NETSV\$D\$MK\$ND\$PMOY\$YASIAL\$NA\$VM\$IL\$KA\$IG\$V\$V\$K\$Q\$TL\$R\$AS\$RL\$H\$DE\$F\$RL | 956 |
| Db | 1001 | NNT\$V\$-----RLGY\$ATL-GILQGLVIML\$AFIMV\$--GAIDARLLTHALLHQI | 1048 |
| QY | 957 | R\$PK\$F\$EDT\$T\$TG\$ILNR\$F\$K\$D\$E\$D\$V\$RL\$P\$Q\$A\$E\$F\$ION\$V\$LV\$F-----F\$Y\$G\$M\$G | 1009 |
| Db | 1049 | RAPO\$F\$E\$PT\$P\$SG\$RLNR\$F\$SDI\$V\$IH-----EVLAP\$ILM\$F\$N\$F\$Y\$T\$S\$T\$V\$V\$A | 1101 |
| QY | 1010 | V\$P\$M\$V\$AV\$G\$V\$LL\$F\$V\$M\$H\$V\$S\$V\$IL\$REL\$KL\$D\$NT\$Q\$P\$LS\$HT\$S\$T\$O\$G\$AT\$J\$H\$AV\$K | 1065 |
| Db | 1102 | STP\$E\$CV\$V\$LE\$A\$F\$Y\$G\$V\$OR\$F\$V\$V\$AT\$S\$R\$Q\$LR\$E\$S\$R\$S\$T\$P\$S\$H\$E\$S\$E\$Y\$T\$G\$H\$V\$IR\$A\$G\$R | 1166 |
| QY | 1070 | GQ\$E\$FL\$H\$V\$O\$E\$LL\$D\$N\$Q\$AP\$F\$E\$TC\$AM\$R\$W\$LA\$V\$RL\$D\$LSIALITTTG\$M\$IV\$M\$G\$Q\$IP\$P\$A\$V\$A | 1125 |
| Db | 1162 | VQD\$F\$V\$V\$LD\$A\$V\$D\$N\$Q\$M\$T\$T\$P\$Y\$A\$S\$N\$M\$W\$G\$V\$H\$E\$P\$G\$N\$CV\$LE\$F\$A\$F\$V\$IG\$N\$S\$N\$P\$G\$LV | 1222 |
| QY | 1130 | GLA\$Y\$AV\$Q\$LT\$G\$LF\$Q\$TV\$K\$A\$S\$E\$T\$E\$A\$F\$T\$S\$E\$V\$E\$IN\$H\$IK\$T\$S\$LE\$A\$P\$ARK--NK\$AP\$D\$P\$M | 1188 |
| Db | 1222 | GL\$Y\$V\$A\$LO\$V\$T\$S\$LM\$M\$M\$IR\$E\$LD\$E\$S\$NI\$V\$E\$V\$E\$K\$E\$S\$K\$T--E\$T\$E\$A\$P\$W\$E\$S\$N\$R\$A\$E--G\$M\$P | 1279 |
| QY | 1189 | Q\$E\$V\$T\$E\$P\$N\$A\$M\$R\$E\$M\$P\$LY\$IK\$V\$ST\$T\$PK\$K\$IG\$V\$G\$T\$G\$S\$K\$S\$LM\$A\$P\$R\$V\$E\$L\$S\$G | 1248 |
| Db | 1280 | R\$G\$V\$E\$E\$F\$N\$S\$V\$R\$P\$G\$E\$LV\$L\$K\$N\$L\$TH\$V\$G\$E\$K\$V\$G\$V\$G\$T\$G\$A\$G\$S\$M\$K\$LC\$F\$R\$L\$E\$A\$E | 1333 |
| QY | 1249 | GC\$IK\$ID\$V\$R\$IS\$D\$IG\$AD\$R\$SK\$LS\$T\$I\$D\$E\$P\$LF\$S\$G\$Y\$R\$S\$N\$D\$P\$E\$N\$Q\$Y\$D\$O\$M\$A\$D\$E\$R\$TH\$M | 1308 |
| Db | 1340 | GE\$IF\$ID\$G\$N\$V\$A\$H\$IG\$H\$D\$LR\$Q\$LT\$T\$I\$P\$D\$P\$IL\$F\$S\$G\$T\$R\$M\$N\$D\$P\$G\$R\$S\$D\$E\$D\$M\$D\$W\$R\$E\$LS\$H\$L | 1399 |
| QY | 1309 | K\$C\$T\$A\$Q\$P\$L\$K\$E\$S\$E\$W\$E\$N\$D\$F\$S\$Y\$G\$E\$Q\$LIC\$A\$R\$A\$LL\$R\$CK\$I\$LL\$D\$E\$A\$T\$A\$D\$E\$T\$D\$LI | 1366 |
| Db | 1400 | SA\$F\$V\$S\$O\$P\$T\$G\$ID\$F\$O\$C\$S\$G\$C\$D\$N\$L\$S\$V\$G\$O\$L\$V\$C\$A\$R\$A\$LL\$R\$K\$S\$R\$V\$LY\$V\$D\$E\$A\$A\$A\$D\$LE\$T\$D\$LI | 1459 |
| QY | 1369 | Q\$E\$T\$R\$E\$A\$D\$Q\$T\$M\$T\$LI\$A\$H\$R\$H\$V\$G\$S\$R\$IV\$N\$A\$O\$G\$V\$E\$D\$P\$P\$V\$S\$L\$M\$D\$S\$R\$E\$F\$A\$M | 1425 |
| Db | 1460 | Q\$G\$T\$R\$Q\$E\$D\$C\$T\$V\$T\$LI\$A\$H\$R\$LT\$M\$D\$W\$K\$R\$V\$LY\$D\$K\$G\$V\$A\$E\$D\$S\$V\$N\$L\$A--AG\$IF\$IG\$M | 1515 |

| | | | |
|----|------|---|------|
| Qy | 1010 | VEFMIVAVGVLVILFESVIMHIVSVILIREKRLRDINTOSPELSHTSSIQGLATTHAVNK | 1065 |
| Db | 1102 | STPLECVVLELAEVFEYFVORFVAVTSRÖLKRLESYRSRSDTSHSEFEYVGTGHSVIRAGR | 1167 |
| Qy | 1070 | GOEFHLRYOELLDDNOAPFELTFCAMRWLAVRLDLISIALITTTGMIYIMHGQIPPAYA | 1129 |
| Db | 1162 | VQDFEVLSDADVDSNÖMTTPYTIASNNWLGVHFEFGNCVYLESAFPAVIGNSNLPGVY | 1227 |
| Qy | 1130 | GLAISYAVQLTGLFQFTVRLASETEARFETSVERIINHITKLSLEAPARIK-NKAPSPDP | 1188 |
| Db | 1222 | GLSYSAVQVLTSLNMMIRITLSDLESNIIVAEVYKESKRT-EFEAPWYLESNRABE-GMP | 1279 |
| Qy | 1189 | QGEYTFENAMARVREMLPIYIKVSTTIPKPKKIGVGTSGSGKSLGMALPRLVELSG | 1248 |
| Db | 1280 | RSGVFEFYNYSVRYRPGLEIVLELKNLTLHVGVGKRVGVGTGAGKSSMVLCLFRILEANE | 1333 |
| Qy | 1249 | GCIKIDGRISIDIGLADLRSKLSITPOEPLFEGYATSNLDPEFNÖYTEDOIMDALERTHM | 1308 |
| Db | 1340 | GEIFLDGNVAHIGLHDLRSQTLTIIPQDPIFLFSGTLRMLNDPGRVSDDEDIWRTELSHL | 1399 |
| Qy | 1309 | KECIAQPLFKLSEVMENGDNFSYGERÖQLICARALLIRHCKIILIDEATAMDETDLII | 1366 |
| Db | 1400 | SAFVSQPTGLDFQCSGEGCNLSVGQOLCVCLARALKRSRYVLVDEAFAAIDLETDLII | 1459 |
| Qy | 1369 | QETIRAEADQETMLTIHARHTVAGSRIYVLAQOQVFEEDPVSYLSMDSSRFYAM | 1425 |
| Db | 1460 | QGTIRTOGEDCTVTLIARHRLNTIMDKRYVLVDKGVAAEDSSVNLTA-AGSIFEGM | 1515 |

RESULT 11
MRP3_HUMAN STANDARD; PRT: 1527 AA.
ID MRP3_HUMAN 060265; 0609222; 075621; 095078; 095289; 095290;
AC 015438; 060265; 0609222; 075621; 095078; 095289; 095290;
DT 15-JUN-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE CANALICULAR MULTISPECIFIC ORGANIC ANION TRANSPORTER 2 (MULTIDRUG
DE RESISTANCE-ASSOCIATED PROTEIN 3).
OS ABC3 OR CMOAT2 OR MRP3 OR MRP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN RP
RP SEQUENCE FROM N.A.
RX MEDLINE=98409330; PubMed=9738950;
RA Kluich Y., Suzuki H., Hirohashi T., Tyson C.A., Sugiyama Y.;
RT "CDNA cloning and inducible expression of human multidrug resistance
RT associated protein 3 (MRP3).";
RL FEBS Lett. 433:149-152(1998).
RN RN
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Keppeler D., Koenig J.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN RN
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99032812; PubMed=9813153;
RA Uchilumi T., Hinooshita E., Haga S., Nakamura T., Tanaka T., Toh S.,
RA Furukawa M., Kawabe T., Wada M., Kagotani K., Okumura K., Kohno K.,
RA Akiyama S., Kuwano M.;
RT "Isolation of a novel human canalicular multispecific organic anion
RT transporter, cMOAT2/MRP3, and its expression in cisplatin-resistant
RT cancer cells with decreased ATP-dependent drug transport.";
RL Biochem. Biophys. Res. Commun. 252:103-110(1998).
RN RN
RN SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kool M., de Haas M., Ponne N.J., Baas F., Borst P.;
RT "Complete coding sequence of human MRP3, a homolog of the human
RT multidrug resistance-associated protein MRP1.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN RN
RN SEQUENCE FROM N.A. (ISOFORMS 3A AND 3B).
RC TISSUE=Liver;
RX MEDLINE=99107643; PubMed=9889399;
RA Fromm M.F., Leake B., Roden D.M., Wilkinson G.R., Kim R.B.;
RT "Human MRP3 transporter: identification of the 5'-flanking region,
RT genomic organization and alternative splice variants.";
RL Biochim. Biophys. Acta 1415:369-374(1999).
RN RN
RN SEQUENCE OF 1043-1527 FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=97413640; PubMed=9270026;
RA Kool M., de Haas M., Scheffer G.L., Scheper R.J., van Eijk M.J.,
RA Juin J.A., Baas F., Borst P.;
RT "Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5,
RT homologues of the multidrug resistance-associated protein gene
RT (MRP1), in human cancer cell lines.";
RL Cancer Res. 57:3537-3547(1997).
CC -1- FUNCTION: MAY ACT AS AN INDUCIBLE TRANSPORTER IN THE BILIARY AND
CC INTESTINAL EXCRETION OF ORGANIC ANIONS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 3 (SHOWN HERE), 3A AND
CC 3B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE LIVER. ALSO EXPRESSED
CC IN SMALL INTESTINE, COLON, PROSTATE, TESTIS AND BRAIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AB010867; BAA28146.1; -
DR EMBL; Y17151; CAAT6658.1; -
DR EMBL; AF083552; AAC34668.1; -
DR EMBL; AF009670; AAD01430.1; -
DR EMBL; AF085690; AAD02845.1; -
DR EMBL; AF085691; AAD02846.1; -
DR EMBL; AF085692; AAD02847.1; -
DR EMBL; U83659; AAB71756.1; -
DR HSSP; P13569; INBD.
DR MIM; 604323; -
DR InterPro; IPR001140; -
DR InterPro; IPR001617; -
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00605; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Alternative splicing.
FT 1 32
FT DOMAIN 1 32
FT TRANSSEM 33 53
FT DOMAIN 54 73
FT TRANSSEM 74 94
FT DOMAIN 95 99
FT TRANSSEM 100 120
FT DOMAIN 121 132
FT TRANSSEM 133 153
FT DOMAIN 154 171
FT TRANSSEM 172 192
FT DOMAIN 193 302
FT TRANSSEM 303 323
FT DOMAIN 324 349
FT TRANSSEM 350 370
FT DOMAIN 371 426
FT TRANSSEM 427 447
FT DOMAIN 448 450
FT TRANSSEM 451 471
FT DOMAIN 472 533
FT TRANSSEM 534 554
FT DOMAIN 555 576
FT TRANSSEM 577 597
FT DOMAIN 598 964
FT TRANSSEM 964 984
FT DOMAIN 985 1021
FT TRANSSEM 1022 1042
FT DOMAIN 1043 1085
FT TRANSSEM 1086 1106
FT DOMAIN 1107 1107
FT TRANSSEM 1108 1128
FT DOMAIN 1129 1199
FT TRANSSEM 1200 1220
FT DOMAIN 1221 1222
FT TRANSSEM 1223 1243
FT DOMAIN 1244 1527
FT NP_BIND 661 668
FT NP_BIND 1323 1330
FT CARBOHYD 18 1006
FT CARBOHYD 1007 1007
FT VASPLIC 226 510

FT EXTRACELLULAR (BY SIMILARITY).
FT 1 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 2 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 3 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 4 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 5 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 6 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 7 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 8 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 9 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 10 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 11 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 12 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 13 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 14 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 15 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT 16 (BY SIMILARITY).
FT EXTRACELLULAR (BY SIMILARITY).
FT 17 (BY SIMILARITY).
FT CYTOPLASMIC (BY SIMILARITY).
FT ATP (POTENTIAL).
FT ATP (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MAITGSRHPLEKRLMSLKEDRSOMVVOOLELAEKROEKO
FT TARRKASABPKNAGSDVILGARRPPRRSPKALIAATP
FT GSSFLISACFKLIDDLISFINPQLSLIRFISNPAPSMW
FT GFLVAGLMFLCSMQSLILQHYHYIVTGKFTGIMGYL
FT YRKALVITNSYKRASTGEIYVNLMSVAPMLDAPLNLV
FT WSPAPLITAIYFLMONGPVLAVGAVPMVLLPGLGAVAV
FT KMRAPFOVKMKLRDSIRKIMSEILINGIKVILYAMPSF
FT -> LINDPLNGCLPGFISPDQGHMLPASGSGSLPVKGS

```
FT  RGQIPDGGAAAAAGMEAGKADGTTGCFSTWCKLRGRG
FT  AAGCPAOPREALPPEGPAGHLRLPHOCLLOAYGPPALLH
FT  OSTAQHDPDOYVLOPHGHLGLOLPGMADVPYLHNAVADLT
FT  TLPLHLCDEWEVSTWDCGHLQDSGSGHQUSQVCHGCGN
FT  COPHVSAGCPALHPCPLQASAVSPADHPGDLPLAPRS
FT  LCPGMSRPHGLADSTORSCGREDAHLPGKANEIEGLADQAD
FT  E (IN ISOFORM 3B).
FT  MISSING (IN ISOFORM 3B).
FT  VARSPLIC 511 1527
FT  VARSPLIC 1194 1527
FT  WLSGCEFGNVCYVFPALFANVIGSSINLPGVLGVSYSKL
FT  OYTFALNMIRMSDLESINIAVERKEYSKETELAPWVE
FT  GSRPEGMPPRGVEFRNYSVRPGLDVLNIDSLHVG
FT  EKVGIYKRTGAGKSSMTLCFLRIEAAAGEIIDLNVADI
FT  GLHDLRSQITIPDPIFSGTILRMLNLPFGSSEEDIMWA
FT  LELSHLTFVSSOPAGLDPGCGEGENISVCGROLYCLARA
FT  LLKRSRIIVDBATAIDLETDNLQATIRTOFDCTYIAT
FT  AHRNTIMDYTRVLDGVAAEDVSPALILAKRIFIGMA
FT  RDAGLA -> SEASALPCCSSRNQOALMCSLSLSPKO
FT  KTGAPLPHPLLI (IN ISOFORM 3A).
FT  C -> R (IN REF. 2).
FT  FHGLVHGR -> SMASMA (IN REF. 2).
FT  A -> T (IN REF. 5).
FT  A -> G (IN REF. 3).
FT  MELCSM -> DVPAVAP (IN REF. 4).
FT  F -> Y (IN REF. 1).
FT  F -> C (IN REF. 6).
FT  L -> I (IN REF. 6).
FT  D -> E (IN REF. 6).
FT  CLF -> ACS (IN REF. 6).
FT  L -> F (IN REF. 6).
FT  LRSOL -> VRCOM (IN REF. 6).
FT  L -> V (IN REF. 3).
FT  L -> R (IN REF. 6).
FT  Q -> R (IN REF. 6).
FT  SEQUENCE 1527 AA; 169342 MW; 0D1F879B6F18370C CRC64;
```

Query Match 26.6%; Score 1945; DB 1; Length 1527;
Best Local Similarity 32.3%; Pred. No. 1.2e-118;
Matches 464; Conservative 268; Mismatches 510; Indels 196; Gaps 22;

```
QY 85 LSAI-----KPIRTCK-----HOHPVDNAGLFSCMTFSWLSLARVAHKKGELSMEDV 133
DB 182 LSAIACREKRPPEFSAKVNDPNPYERSAGFLSRLFFWMFTKMAIYGRH-PLERKDL 240
QY 134 WLSKHESSDVACRLERLMQELNEVGPDA-----L 167
DB 241 WLSKEDRSQMVVODLEAMRKQKOTARHKASAAFGKNAEGEDVLLGARPRKRPSEFL 300
QY 168 RRVVWIFCETRLILSIYCLMITQIAGFSGAPAFMKHLLLEYQATESNLQYSLILVGLLL 227
DB 301 KALLATF-GSSFLISACFKLIQDLSFINPO--LSTILIRFISNPMAPSWMGFLVAGLMLFL 358
QY 228 TEIVSWSLATWALNTRTVGRVLRGAILTMAFKKILKLN-IKES-LGELINICSDQO 285
DB 359 GCMQSLILQHTYHYTFVGVKFRGIMGVIRKALVITNSKRASVGEIYVLMVSDQO 418
QY 286 RNFEEAANGSLLAGGVVAAILGMIIVIIILPPTGLGSAVFILFPAMMPASRLTAFFRR 345
DB 419 RFPMDLAPFLNLTMSAPLQIILAIYFLMOMGPGSVLAGVAFMVLILPLNGAVVAKMRAFOV 478
QY 346 KCVAAATDEVOKANEVLYITIKFIKMYAWKAPSOYOKIRREERRIRLEAGYFQSTIVGV 405
DB 479 KQMKLKDSIKILMSSELNLIKVKLKYAMPSPFLKQVEGIRQGLLRLKRYAAVYHTTTPPT 538
QY 406 ---APIVVVIAVSVTFVSHMTLGFDTLTAQAFTVVTVFNSMTFALKVTFPSVKSLSSEASY 462
DB 539 WMCSPFLVTLITLIMVY-VYVDNNVLDAAEKAFYSVSLFILRLPLMLLQLLSNLQASV 597
QY 463 AYDRKSLFLAMEV--HMINKRPASPHIKIEKKNATLANDSSHSSIQNSPKLTPKAKKOK 520
DB 598 SLKRIQOFLQSELDPOASVERKRTISPGAITITHSCTFAOD-----639
QY 521 RASRGKKEKVRLOQRTHEQAVLAEGKHLLDSDERPSPEEERKHIHGLHLQRTLHS 580
```

```
DB 640 -----LPETLHS 646
QY 581 IDLEIOEGVLGICVSGSKTSLISAILGQMTLEGGIAISGTFAYVAQOAMILNATR 640
DB 647 LDIQVKGALVAVVGVGGKSSLSVALGEMERKEGKHYMMGSVAYVQOAMIOQCTQO 706
QY 641 DNILFEKDEDEREYNSVLNSCLRPDLAIIPSSDLTEIEGERANISGQORISILARALY 700
DB 707 ENVLKSKALNPKRYQOTLACALIDLLEMLPGDDQTEIEGERKINLSGQORQISILARAY 766
QY 701 SDRSIYIIDDPLSALDAVGHNIENSARKH--LNSKTVLFYTHQLOLYLVCDEVEIFMKE 758
DB 767 SDADIFLDDPLSAVDSVSHAKHIFDHVIGPEGLAGKTRVLTWHSIFPQDFEITVLAD 826
QY 759 GCITTEGTHIEELMNLNGDYA-----TIFNN 783
DB 827 GQVSEMGPPALLQRRNGSFANFLCNYPAPDEDOGHLEDSTALEGADKREALLIEDTLSSH 886
QY 784 LLLGETPPV-----EINSKETSQSKQSKQPKPTGSIKREKAVKPEEGOL 830
DB 887 TDLTNDPVTYVYQKQPMQOLSNLSDGCGRPVRRRLGP---SEKQVTEAKADGAL 943
QY 831 VOLEEKQGSVPMSYVYVYIOAAGPLAFLVJMALFMLNVGSTAF--STWMLSYWIKOG 887
DB 944 TQEKKAIGTVELSVWDYAKAVG---LCTTLAICLLYVGQSAALIGANVWLSAW-----995
QY 888 SCNTYVTRGNETSVSDSMKDNPHMQYTAITYALSMANVLLKAIKRGVAVKGLRASSSL 947
DB 996 -----TNDAMADSRONNTSLR--LGVTYALGILQGLFVWLAAMAAAGIOQARVL 1044
QY 948 HDLEFRILRISPMKFPDPTPTGSLIRFSKDNDEVDVRLPEQAEPTOMVILVFCVGM 1007
DB 1045 HQALHNKIRSPQSFDTYPSGRILNCSKDIYVDEVALPYLMLNLSFNNAISLTVYI 1104
QY 1008 AGVFPWFLVAVGPVLIFFLVHIVSRVLRLEKRLDNIQOSPFLSHITSSIOGIATIHAY 1067
DB 1105 MASTPLFTVYVILPLAVLYTYQRFYAATSRQLRLESVERSPYSHFSEVGVASVYIRAY 1164
QY 1068 NKGOEFLHRYOELLDNQAPFLFTCAMRWLAVRLDLISALITTTGMIYLMHQIIPPA 1127
DB 1165 NRSRDEIISDRKVDANQSCPYIISNRMLSIGVEFVNCVVLFPALRAVIGRSSLNG 1224
QY 1128 YAGLAISYAVOLTGQFVYVRLASETEARFTSYERNHVIKTLISLAPARIKKAPSPDM 1187
DB 1225 LVGLSVSYSLQVTFALNMWIRMSDLESINIYAVERKEYSKT-ETPAVAVESGRPECG 1283
QY 1188 PQGEVTFENAEKRYRENDPLYLKVSFTIKPEKIGIYGRSGKSSGLMFLRVELS 1247
DB 1284 PRGGEYEFERNYSVRYPGLDYLRDLSLHVHGEKGYIGRICGAGSSMTLCFLRIEAA 1343
QY 1248 GGCIRKIDGVRISDIGLADLRSLIIPQEPVLESFGTVRSNLDPEFNOYTEDQIWDALERTH 1307
DB 1344 KGEIRIDGLNADIGLHDLRSQITIIPOPILESGLTRMLNLPFGSYSEBEDIWMALELSH 1403
QY 1308 MKECTAQDLPLESEVEWNGDNFVSGBEROLLCAARLLHRCKLIIIDETAADTDTDLL 1367
DB 1404 LHPFVSSOPAGLDPOCSGEGENISVGOQLVCLARALLKRSRLVLDXETAALDEFTDM 1463
QY 1368 IOETIEAFAOCTMLTIFARLHTVLGSDRIMVLAQOQVEFDPSPVILLSNDSSRFAM 1425
DB 1464 IQATITGOTDCTCVLTIAIRLNTIMDYTRVLYDKQVAAEFEDSPANLIA-ARQIFIGM 1520
```

RESULT 12
YAMB.SCHPO
ID YAMB.SCHPO STANDARD; PRT: 1478 AA.
AC 010185;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROBABLE ATP-DEPENDENT PERMEASE C3F10.11c.
GN SPAC3F10.11c.
OS Schizosaccharomyces pombe (Fission yeast).


```

Db 1102 SIGGLSTIRAYMEDFEISENDIRVDNHRHWFELFYSSNRMOAIRVEAIGALVWSSAF 1161
Oy 1117 IYL--MHQOIPRANGLASVAVOLTGLFQFVVRCLASEFAFTSVERNHKLISLEA 1174
Db 1162 GLSARVGNPSGLVLSLSTAVQITQSLTFVVRQSVDEINIVSERMLEYI-GLPSRA 1220
Oy 1175 PARINKKAPSPMPOEGEVTFENAMRRYENPLVLKKKFTIKPEKIGIARTSGSKS 1234
Db 1221 PSTIDHRPREMPSHGAIKPFHYSVRRENPLVINDISVNIKQEKIGIARTSGAGKS 1280
Oy 1235 SIGMALFRLVELSGGCIKIDGVRSIDIGLADLRSLIIPQEPVLFGSTVRSLDPFNQY 1294
Db 1281 TLTFLFRLIETPSGDIQDIDINITSIGLHDSRLAIIPQENQAFEGTIRRENDPNANA 1340
Oy 1295 TEDQWDLERHMECAQPLKLESEYMGNDNFVGEROLLCIARLLHCKKLIID 1354
Db 1341 TDEEWHLEASLQFOTLDGGYSRTEGGANLSSQORLMCTRLRLPTRVLLD 1400
Oy 1355 EATAMDETDLIQETIREAFADCTMLTIAHRLHVLGSDRIMVLAQGVVEFDPSPVL 1414
Db 1401 EATAVDVETDAIVQRTIRENDRTIILTAHRTVADSNRLVLDHCKVVEPDSKTL 1460
Oy 1415 LSNDSSREFYAM 1425
Db 1461 LENKASLEFYSL 1471

RESULT 13
ACCG_HUMAN STANDARD; PRT; 1580 AA.
AC 009428; 075948; 016583;
DT 01-NOV-1995 (Rel. 32, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SULFONYLUREA RECEPTOR 1.
GN ABC8 OR SUR1 OR SUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBL_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gonzalez G., Aguilar-Bryan L., Bryan J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE-Pancreas;
RA Nishimura M., Miki T., Alzawa T., Seino S.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA TISSUE-Foreskin;
RA Thomas P.T., Wohlik N., Huang E., Gagel R.F., Cote G.J.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1186-1580 FROM N.A.
RA TISSUE-Pancreatic islets;
RA MEDLINE-95232533; PubMed=7716548;
RA Thomas P.M., Cote G.J., Wohlik N., Haddad B., Mathew P.M., Rabi W.,
RA Aguilar-Bryan L., Gagel R.F., Bryan J.;
RT "Mutations in the sulfonylurea receptor gene in familial persistent
RT hyperinsulinemic hypoglycemia of infancy.";
RL Science 268:426-429(1995).
RN [5]
RP TOPOLOGY.
RP MEDLINE-99436117; PubMed=10506167;
RA Raab-Graham K.F., Cirillo L.J., Boettcher A.A., Radeke C.M.,
RA Vandenbergh C.A.;
RT "Membrane topology of the amino-terminal region of the sulfonylurea
RT receptor.";
RL J. Biol. Chem. 274:29122-29129(1999).
RP [6]
RP REVIEW ON VARIANTS.

RX MEDLINE-99268411; PubMed=10338089;
RA Meisner T., Beilbrech B., Mayatepek E.;
RT "Congenital hyperinsulinism: molecular basis of a heterogeneous
RT disease.";
RL Hum. Mutat. 13:351-361(1999).
RN [7]
RP VARIANT PPHI VAL-715.
RX MEDLINE-96354544; PubMed=8751851;
RA Thomas P.M., Wohlik N., Huang E., Kuhnle U., Rabi W., Gagel R.F.,
RA Cote G.J.;
RT "Inactivation of the first nucleotide-binding fold of the
RT sulfonylurea receptor, and familial persistent hyperinsulinemic
RT hypoglycemia of infancy.";
RL Am. J. Hum. Genet. 59:510-518(1996).
RN [8]
RP VARIANT ALA-1368.
RX MEDLINE-96220088; PubMed=8635661;
RA Inoue H., Ferrer J., Welling C.M., Elbein S.C., Hoffman M.,
RA Mayorga R., Warren-Perry M., Zhang Y., Millins H., Turner R.,
RA Province M., Bryan J., Permutt M.A., Aguilar-Bryan L.;
RT "Sequence variants in the sulfonylurea receptor (SUR) gene are
RT associated with NIDDM in Caucasians.";
RL Diabetes 45:825-831(1996).
RN [9]
RP VARIANT PPHI PHE-1386 DEL, AND VARIANTS GLY-1359; ALA-1368 & ILE-1571.
RX MEDLINE-97081768; PubMed=8923011;
RA Nestorowicz A., Wilson B.A., Schoor K.P., Inoue H., Glaser B.,
RA Landau H., Stanley C.A., Thornton P.S., Clement J.P. IV, Bryan J.,
RA Aguilar-Bryan L., Permutt M.A.;
RT "Mutations in the sulfonylurea receptor gene are associated with
RT familial hyperinsulinism in Ashkenazi Jews.";
RL Hum. Mol. Genet. 5:1813-1822(1996).
RN [10]
RP VARIANT PPHI ARG-1477.
RX MEDLINE-96256476; PubMed=8650576;
RA Nichols C.G., Shyng S.-L., Nestorowicz A., Glaser B., Clement J.P. IV,
RA Gonzalez G., Aguilar-Bryan L., Permutt M.A., Bryan J.;
RT "Adenosine diphosphate as an intracellular regulator of insulin
RT secretion.";
RL Science 272:1785-1787(1996).
RN [11]
RP VARIANTS GLN-274; MET-559; ASN-809; CYS-833 AND ALA-1368.
RX MEDLINE-98178572; PubMed=9519757;
RA Ohta Y., Tanizawa Y., Inoue H., Hosaka T., Ueda K., Matsutani A.,
RA Repunte V.P., Yamada M., Kurachi Y., Bryan J., Aguilar-Bryan L.,
RA Permutt M.A., Oka Y.;
RT "Identification and functional analysis of sulfonylurea receptor 1
RT variants in Japanese patients with NIDDM.";
RL Diabetes 47:476-481(1998).
RN [12]
RP VARIANTS PPHI
RX MEDLINE-98282238; PubMed=9618169;
RA Nestorowicz A., Glaser B., Wilson B.A., Shyng S.-L., Nichols C.G.,
RA Stanley C.A., Thornton P.S., Permutt M.A.;
RT "Genetic heterogeneity in familial hyperinsulinism.";
RL Hum. Mol. Genet. 7:1119-1128(1998).
RN [13]
RP VARIANTS PPHI PRO-1351; CYS-1419 AND TRP-1492.
RX MEDLINE-98443218; PubMed=9769320;
RA Verkarre V., Fournier J.-C., de Lonlay P., Gross-Morand M.-S.,
RA Devilliers M., Rahier J., Brunelle F., Robert J.-J., Nihoul-Fekete C.,
RA Saudubray J.-M., Junien C.;
RT "Paternal mutation of the sulfonylurea receptor (SUR1) gene and
RT maternal loss of 11p15 imprinted genes lead to persistent
RT hyperinsulinism in focal adenomatous hyperplasia.";
RL J. Clin. Invest. 102:1286-1291(1998).
RN [14]
RP VARIANT PPHI ASP-186.
RX MEDLINE-99265491; PubMed=10334322;
RA Otonokoshi T., Aemmelaelae C., Huopio H., Cote G.J., Chapman J.,
RA Cosgrove K., Ashfield R., Huang E., Komulainen J., Ashcroft F.M.,
RA Dunne M.J., Kere J., Thomas P.M.;
RT "A point mutation inactivating the sulfonylurea receptor causes the

```



```

Db 1022 HMYLVAD-YMLAKWDSALTLTPAAR---NCSLSQECTLDQTVYAMVTVLCSIGIVL 1076
QY 929 KAIRGVVFKGLRASSRIHDELFRILRSPKFFDTPTGILNFSQDMQEDVRLPF 988
Db 1077 CLVSTVEWMTGLKAKRLHSLRLRIILAPMRFFETPLGSLINFSQSDCTIDQHIPS 1136
QY 969 QAEMLIONVILVFCVGMAGVPMFVLAVGPIVLFSVLHIVSRLLINELRLDINTQS 1048
Db 1137 TLECLSRVLLCVALAISTVYVFLVALPLAIVCYIQKFRVASHDDLDPTDQL 1196
QY 1049 PELSHITSSIOGLATIHAYNKQGEFLHRYOELDDNOAPFFLETCAMRMVAYRLDLIS-- 1106
Db 1197 PLLSHAFVEGLVLTITRAFERYARFOQKLEVTDSNNIASLEFLTANRLLEVMEITGAC 1256
QY 1107 IALITTTGLMVLHGOIPPAVAGLAISTAVOLTGLFOFTVRLASTEARFTSVVERINHX 1166
Db 1257 VVLLAAVTSISLSHRELHAGLVGLGTALYALVSNYLMNMVNRNLADMEQLQAVKRHL 1316
QY 1167 IKTLSEAPARIKNKAPS---PDWPOGEVTFENAMERYREMLPYLKKVSFTIKPKKI 1223
Db 1317 LKT---EASVYGLAPSLIPKMPDQKIQIONSVRSDSLKPKVLKHNALISFGKI 1373
QY 1224 GIVGRSGKSSLSGALFELVELSGGCIKIDGVRISDIGLADLRSLIIPQEPVLSGT 1283
Db 1374 GICGRSGSKSSFLAFPMVDTPEGHITIDGIDIAKLPLHLRLSLITLDDPVLSGT 1433
QY 1284 VASNDPFRQYEDQIDALETHKECTAOLPLKIESVWENGDNFSVGERQLLCIARA 1343
Db 1434 IFENIDPERKCSDSLWEALTAOLKLVKALPGGIDALITTEGENFSGOGRQLCIARA 1493
QY 1344 LRHCKILIDETADMETDILQETIREAFADCTMTLTIARHRTVIGSRINWLAOG 1403
Db 1494 FVRKTSIFLMDATASIDMATENILQKVMTAFADRTVITAHRTILSADLVYLVKRG 1553
QY 1404 QVEFDPYSVLLSNDSPRYAFMAAE 1430
Db 1554 ALLEFDKPEKLSRKDSVF-ASFVRAD 1579

RESULT 14
YHDS_YEAST STANDARD; PRT; 1592 AA.
AC P38735;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN PROBABLE ATP-DEPENDENT PERMEASE YHL035C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faveille A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kueba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latteille P., Louis E.J., Macri C., Mardis C., Meneses S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevasaks E., Vaughan K.,
RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA Vaudin M.:
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VII.1."
RL Sequence 265:2077-2082(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: U1583; AB65047.1; -.
DR PIR: SA8933; S48933.
DR HSSP: P13569; INBD.
DR SGD: S0001027; YHL035C.
DR InterPro: IPR001140; -.
DR DR Pfam: PF00664; ABC membrane; 2.
DR DR Pfam: PF00005; ABC trans; 2.
DR DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KW Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
KW Transport.
FT DOMAIN 1 33
FT TRANSMEM 34 54
FT DOMAIN 55 74
FT TRANSMEM 75 95
FT DOMAIN 96 100
FT TRANSMEM 101 121
FT DOMAIN 122 131
FT TRANSMEM 132 152
FT DOMAIN 153 170
FT TRANSMEM 171 191
FT DOMAIN 192 329
FT TRANSMEM 330 350
FT DOMAIN 351 379
FT TRANSMEM 380 400
FT DOMAIN 401 465
FT TRANSMEM 466 486
FT DOMAIN 487 489
FT TRANSMEM 490 510
FT DOMAIN 511 572
FT TRANSMEM 573 593
FT DOMAIN 594 614
FT TRANSMEM 615 635
FT DOMAIN 636 989
FT TRANSMEM 990 1010
FT DOMAIN 1011 1051
FT TRANSMEM 1052 1072
FT DOMAIN 1073 1115
FT TRANSMEM 1116 1136
FT DOMAIN 1137 1137
FT TRANSMEM 1138 1158
FT DOMAIN 1159 1229
FT TRANSMEM 1230 1250
FT DOMAIN 1251 1252
FT TRANSMEM 1253 1273
FT DOMAIN 1274 1592
FT NP_BIND 702 709
FT NP_BIND 1357 1364
FT NP_BIND 1592 AA; 180924 MW; 055FB039992AC68 CRC64;
SQ SEQUENCE

Query Match 23.9%; Score 1749; DB 1; Length 1592;
Best Local Similarity 30.3%; Pred. No. 7.5e-106;
Matches 439; Conservative 288; Mismatches 195; Indels 228; Gaps 33;

QY 86 SALKPRTTC-----KHQHPVDN--AGFSCQMTSWLSLARVAHKKGELSM 130
Db 239 STLDPIKLTSPKDNSSITTYVRDHTSPREHISLSLCTIWSWITNFWEA QKWTIKL 297
QY 131 EDVWSLSKSHSSDVNCRRLERLMOEELNEVGPDASLRVWVFCRTILSLVCLMTIQ 190
Db 298 KIMIGLSMEDYSIFILKGFTRR-NKHIN-----NLTLAFESFTYLLIGMLAVLVNS 349
QY 191 LAGFSPARVWVHLEYT---QATESNLOYSLIVYGLLITELVRSWSLALRYMALRYRG 247
Db 350 IYNLL-PTIMKRFEIYDNPNRSSSCNMJAMLYIIGMFCIRLTALICNSOGQFSDKIC 408
QY 248 VRLRGAILTMRKKIKLK-----NIKESLGELINICSDGQRMFEAA----- 291

```

| | | | |
|----|------|--|------|
| D0 | 409 | LRIRALIGETIYAKGJRRRLFTSPKTSOSDSISANLGTIIMLISIDSPKSEYSELANLYV | 468 |
| QY | 292 | AVGSLIAGRPVVALIGMITYVILGPGLGSAVFILLEYPAMMFASRLTAIRKRCVAAT | 351 |
| D0 | 469 | TVQAVIM---IIYVVGILLEN--FLGSAFAGISIIILYMPPLNFLANLILGKFOKOTILKCT | 523 |
| QY | 352 | DEROKNEVELYTIKFIKIMYAWKAFSOSQOKREERERILEKAGFOSTIV---GVAPI | 408 |
| D0 | 524 | DORISKTNECLQNRITYKFAWERNIINEKTSIROKELSLKLKSLWVSITFLMFVPT | 583 |
| QY | 409 | VV--VIASVVTFSYHMTLGFEDLPAQAQFTVYFENSMTPALKYTFPFSYSLEASVAADR | 466 |
| D0 | 584 | LYGVTAICTPFOHE---DLNAPLAFITLISFTLKTRPDQLSMLSLFINSQVSLKR | 639 |
| QY | 467 | FKSLFLEEYHMIKNKRPASD-HIKIEMKNATLAMDSSHSSIONSPLKAPPKMKDKRASRG | 525 |
| D0 | 640 | ISDLFRDDTPEKYNQULISPDKNKIEFNATLWNNENDSDM--NAFRLC----- | 686 |
| QY | 526 | KKEKVRLOQRTHEQAVLAEDKGHLLDSDERSPEEBEKGHILHGLRLQRLHSIDLEI | 585 |
| D0 | 687 | -----GLNTKF | 692 |
| QY | 586 | QEGKLVGICGSVSGKTSLSALIGOWTLEESI-----AISGFAVYA | 629 |
| D0 | 693 | QIGKLNILIDSTGSGKALLGLIGELNLSGSIYPSLEPKHDLIPDCGLTNSFACIS | 752 |
| QY | 630 | QOAMILNATLRDNLTFEKEXEDERYSVNLSCLRBDLAILPSSDITELTEIGEGANILSGO | 689 |
| D0 | 753 | QSAWLNDYTKNNIIFDNFYNEDRYKKVIDACGLKNDLELPAQDGTETLGEKGITLSGQ | 812 |
| QY | 690 | QORISLARALYSDBRSYIILDDPSALDAHYGNHIFNSAIRKHL--KSYLVLYTHOLOV-L | 747 |
| D0 | 813 | QORISLARAYSSAKHYVLLDDCLSAVDSHTAWIYENCITGPMLKRNCTILTHVNSLTL | 872 |
| QY | 748 | VDOCEVTFMEKGCITBEGTHEELMNLNGDVIATFENLLGEPFPVAINKREKTSOSQKS | 807 |
| D0 | 873 | RNAHFALVLEKGKKNQGITTELQOS--KGLEKERYOL-----SSRDSINEKNAN | 920 |
| QY | 808 | QODKPKGSIKKKEKAVP-----EBGQVQLEKEKQSGVSPMSVGVYVIOAAGPLA | 858 |
| D0 | 921 | RLKRPKRNDQKIEPVIENINFPANPVNDSQGLEEBEKKNGAISPDVYKWLAKFEGS--- | 977 |
| QY | 859 | FLVTALFMLNVGSTAF---STWLSYWIKQSGNVTYRGNE-----TSVSDSM | 905 |
| D0 | 978 | FKALTAFAIYITAOILFISQSWWIRHW---NDTNVRIAPGFAMDLTPLKGMTDSS | 1032 |
| QY | 906 | KDNPHMQYASIALSAWVLLIKATRGVYFVNGTLRASRLHDELFRILISPKKFPDT | 965 |
| D0 | 1033 | KNKINAFYIYLVFLGITIOMLIGERTMTPLSGMRASRKIFNNLLDVLHAQIRFPDV | 1099 |
| QY | 966 | TPTRILINREKSKMDVEDVRLPEQAEFIOMNTVLVFCGMIAAGVPMFLVAVGCVLIFE | 1025 |
| D0 | 1093 | TPVGRIMNRRSKOIEGVDQGLIYLEVITFCILQOASIIFLITVITPRLVYA---VIVE | 1144 |
| QY | 1026 | SVLHIVSRVL---RELKRLDNITQSPFSLHTSSIOGLATTHAANKGOEFLHRYQELID | 1083 |
| D0 | 1150 | VLYFEVQKWLTSARELKRIDSITKSPRIFQHESETLVGCTIRAFEGDEBERFLIEMNNKD | 1203 |
| QY | 1083 | DNOAPEFLFCAMRWLAVRLDLSIALITTTGMLYLMHQGPRAVAGIAISYANQVLGL | 1144 |
| D0 | 1210 | QNNRAPEFLYSYAKWSEFRMDMGAFVLAAGSEFILNLTANIDSGLAGISLYVALFLFDG | 1263 |
| QY | 1143 | FOFVPRFLASETEARFTSEVERINHYITLSEAPARIKKN-----APSDPMQOEEVFYE | 1196 |
| D0 | 1270 | ALMTVLRLYTFEEMMNSVEYLKEX---SLEQENYIGHDEGRILLNESPWKODEFIE | 1322 |
| QY | 1197 | NAEMRYRENTPLVKKVYSFTIKKEKIGVIGRTSGKSSIGALFELVELSGGCIKIDGV | 1256 |
| D0 | 1327 | NLSLRVAPNLPVYIRNYSFKVDQSKIGVIRGAGACKSITITALLEPTIGCIKIDQO | 1386 |
| QY | 1257 | RISDITGLADRSKLSTIPQEPVLFSGTVANSNDPFOYITEDDIMALER----- | 1306 |

```

Db 1387 DISKIDVTLRRSTIIIPQDILFAGRTIKSVNDPYDEDEKIFKALSOQVNLSSHEFEE 1446
Qy 1306 -----THMECTIAQLPLKESEFMWNGDNFSVSEROLLCARALLRHCKILLIDE 1355
Db 1447 VLNSEERENSHNKF-----LNLHTELEAGELNLSQGEROLLEFARSLREPKIILDE 1500
Qy 1356 ATAAAMDTELDLIIETIEAPADCMLEIARLHRTVGSDSRLMVAQGVVEFPDPVLL 1415
Db 1501 ATSSIDYDSDHILQIILNSEFNKSTIIILARLSVLDYDKRIYVADAGEVKEIDRPELL 1560
Qy 1416 SNDSSRFYAM 1425
Db 1561 KDERGIEFYSM 1570

RESULT 15
ACCG_CRICR
ID ACCG_CRICR STANDARD; PRT; 1581 AA.
AC 009427;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SULFONYLUREA RECEPTOR 1.
GN ABC8 OR SUR.
OS Cricetus cricetus (Black-bellied hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_Taxid:10034;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Pancreatic Islets;
RX MEDLINE=95232532; PubMed=7716547;
RA Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV,
RA Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguu K., Bryan J.,
RA Nelson D.A.;
RT "Cloning of the beta cell high-affinity sulfonylurea receptor: a
RT regulator of insulin secretion.";
RT Science 268:423-426(1995).
RN [2]
RP CARBOHYDRATE-LINKAGE SITE ASN-9.
RX MEDLINE=97098100; PubMed=8942641;
RA Nelson D.A., Bryan J., Wechsler S., Clement J.P. IV, Aguilar-Bryan L.;
RX "The high-affinity sulfonylurea receptor: distribution, glycosylation,
RT purification, and immunoprecipitation of two forms from endocrine and
RT neuroendocrine cell lines.";
RL Biochemistry 35:14793-14799(1996).
RN [3]
RP TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITE ASN-9.
RX MEDLINE=99436117; PubMed=10506167;
RA Raab-graham K.F., Cirillo L.J., Boettcher A.A., Radeke C.M.,
RA Vandenbergh C.A.;
RT "Membrane topology of the amino-terminal region of the sulfonylurea
RT receptor.";
RL J. Biol. Chem. 274:29122-29129(1999).
CC -1- FUNCTION: POTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC AND INSULIN RELEASE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). MRP SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: LA0623; AAA99201.1; -
DR HSSP: P13569; INBD.
DR InterPro: IPR000388; -.

```

```
DR InterPro: IPRO00844; -
DR InterPro: IPRO01140; -
DR InterPro: IPRO01617; -
DR Pfam: PF00064; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR PRINTS: PRO1092; SULENYLUREAR.
DR PRINTS: PRO1093; SULENYLURL.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
KV Receptor; Transmembrane; ATP-binding; Phosphorylation; Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 33
FT TRANSMEM 34 54
FT DOMAIN 55 74
FT TRANSMEM 75 95
FT DOMAIN 96 100
FT TRANSMEM 101 121
FT DOMAIN 122 133
FT TRANSMEM 134 153
FT DOMAIN 154 166
FT TRANSMEM 167 193
FT DOMAIN 194 310
FT TRANSMEM 311 330
FT DOMAIN 331 354
FT TRANSMEM 355 375
FT TRANSMEM 376 433
FT TRANSMEM 434 454
FT TRANSMEM 455 457
FT TRANSMEM 458 478
FT TRANSMEM 479 540
FT TRANSMEM 541 561
FT TRANSMEM 562 583
FT TRANSMEM 584 604
FT DOMAIN 605 1004
FT TRANSMEM 1005 1025
FT DOMAIN 1026 1072
FT TRANSMEM 1073 1093
FT TRANSMEM 1094 1137
FT TRANSMEM 1138 1158
FT TRANSMEM 1159 1159
FT TRANSMEM 1160 1180
FT TRANSMEM 1181 1251
FT TRANSMEM 1252 1272
FT TRANSMEM 1273 1276
FT TRANSMEM 1277 1297
FT DOMAIN 1298 1581
FT TRANSMEM 1582 1581
FT CARBOHYD 172 719
FT NP_BIND 1378 1385
FT CARBOHYD 9 9
FT CARBOHYD 1049 1049
FT SEQUENCE 1581 AA; 177015 MW; 0EFCFB3A454E0B1 CRC64;

Query Match 23.8%; Score 1742.5; DB 1; Length 1581;
Best Local Similarity 31.2%; Pred. No. 2e-105;
Matches 450; Conservative 265; Mismatches 527; Indels 199; Gaps 28;

OY 106 LPSCHTFWSLSTLRAHAKKGLMEDWVSLKHESSDVNCRLEKLM-QEELNEVGP 163
DB 1 LSKGTYWMMNFYIKTAHKK-PIDLRALIKLPTMRALNTYQRLCYAFDAQARKDQSPQ 281
OY 223 LSKGTYWMMNFYIKTAHKK-PIDLRALIKLPTMRALNTYQRLCYAFDAQARKDQSPQ 281
DB 1 LSKGTYWMMNFYIKTAHKK-PIDLRALIKLPTMRALNTYQRLCYAFDAQARKDQSPQ 281
OY 164 AASLRVAV-IFCRF---PLTISTVCLMITOLAGFSGPAF---MVKHLLEYTATESNLQ 216
DB 282 GA---RAIRALCHAGGRLLISSTRILLADLIGFAGPCLTIGIVDLKKNHVFQPKTQ 338
OY 217 Y-----SLLLVGLL--TEIVRSWSLALTMALNTYRTGVRALRGAILT 256
DB 339 FLGVYFVSSQEFGLNAYVALVLELALLLQRTPLQASYYVAL-----EGINLRLGAIQT 392
OY 257 MAFKTLIAK-----NIKESLGEINICNSDGRMEAAVAGSLAGPVALIGMTYV 312
DB 393 KYNKIMHSTSNLSMGWTAGOICMLVAIDTQMLMFFELCPNLTMTMPVOIIVGVILLY 452
OY 313 ILGPTGLSAVFIILFYPMMPASRLTAVFRKKVAATDERVQKKNVEVLTYYIKFKMVA 372
```

```
DB 453 YILGVSAALIGAAYIILLAPQYVFAFKLSQAQRTTLESHSNERLKQTNEMLRGKMLIKLYA 512
OY 373 WKAESQSVOKIREERRLLEKAGIFQSTITGVAPLIVVYIASVYTSVAMTL--GEDLTA 430
DB 513 WESIFCSRVVEYFKEMTSIRAFAYTSSISFMNTAIPILAAVITLVGHVSPKESDLS 572
OY 431 AAFVTVYVNSMTFALKYTPFSVKLSSEASVAVDRFKSLFLEEVYMIKRNKASPHIKI 490
DB 573 SVAFASLSLPHILVLPFLISSVYRSTVKALVYQKLSSEASV---IREBCAP----- 625
OY 491 EMKNATLAWDSHSSIONSFKLTPKMKDKRASRGKREKVKQRTORTHOAVLAQRGHL 550
DB 626 -----REPAQOAGKYQAV-----P 641
OY 551 LQSDERPSPPEERKHIHUGHLRQ-----TLHST 581
DB 642 LKVVNRKRPAREEYRDL-LG--PLQRLAPSDGADNFCVOIIGFETWPDGIPPLSLNI 698
OY 582 DLEIQGKLIVGIGSVSGKSTLSIALIGQMTLEGS----- 618
DB 699 TIRIPGQLTMTIVGQVCGKSSLLATLGEMQVSGAVFNWSNLPDSEGDSPSPERETA 758
OY 619 ----IAISGFAYVAQAAMILNATLRDNLIFGKEYDEERYNSVLNSCCLRPDLAII PSSD 674
DB 759 AGSDISRGVAVASQPKMLNTVEENTFESPKNQRYKMYEACSLQPDIDILPHGD 818
OY 675 LTEIGERGANTSGGQORISILARALYSDRSYIILDDPLSALDAHGNHFNISAIRHKL- 733
DB 819 QFQIGERGINSQGORISIVARALYQQTWVFLDPFSLADVHLSLQACILLETLRD 878
OY 734 -SKTVFYTHQDLYVDCDEVITPMKEGCTERGTHERBLMNDYATIFNNLLGEPV 792
DB 879 DKRTVVLVTHKLOLYLPHAMIIAMKQDTIQREGTLDFQSECOLPEHMTKLRNDQDEL 938
OY 793 EIRSKETSGQSKQSDQKQKGTSSIKKEKAVKEPEQVOLVLEKQ----- 838
DB 939 E-----KEYVERKASE---PSQG---LPRMSSRDGILLDEEEBEBAASEEDNLSV 988
OY 839 ----GSVPSVSVGYVQAAGGPIALVIALFMNLVNGSTAFSTWLSYWKQSGNTVT 894
DB 989 LHQRAKIPWRACTRYSSAGIILLSTLVSQILKHMVLA-----IDYWLAKTWDLSALV 1043
OY 895 RGNETSVDQMKPNHMOYVASTAYALSMAVMLLKIRGVVFKGTLRASSRLHDELFR 954
DB 1044 SPARRKCSLSQEDLDQSYVAVFTLLCSGLVLCVTSVTEWGTGLKAKRLHRSILNR 1103
OY 955 ILRSPKFFPTTPTGRILNFRSKDMDEVAVRLPQOAMFTQNYILVFCGMITAGVPWF 1014
DB 1104 IILAPRFETTTPLGSLINFRSSDCNTIDQIHSTLECLSRILLVCYSALTIVSYTPVF 1163
OY 1015 LVAAGPLVILSVLHIVSRVLRILKRLDNIITQSPFLSHITSSIOGLATTIAYNKGQEL 1074
DB 1164 LVALLPLAAVVCYFIQKRYFVAVSRDLQDLDTQLPLVSHAFVEIGLITRAREYARFQ 1223
OY 1075 HRQOELLDNQAPFELFTQCMRWLAVRDLIS--IALITTTGLMIVLMHQOIPRAYGLA 1132
DB 1224 QKLEETDSNNIASLEPLTAANRWLEVCMEYIGACVYVLIATAATISINSLHRELSAGLVLG 1283
OY 1133 ISYAVOUTGLFOFTYRLASFEARFTSVSRINHYIKTSLSEAPARIKRNKPS---PDMQ 1189
DB 1284 LYTALAVSNLNMVNRNLADMEIOLGAVKRIHALKT---EASYSBELLAPSLIRKMPD 1340
OY 1190 EGEVTEENAMRYRENPLVLRKVSTFKIKRKGIVGRTGSGKSSIGMALFRLVELSGG 1249
DB 1341 QGRIOQIONLSVRRDSSILPKLKHVNTLISPGQKIGICGRTSGKSSSLAFPRVAVMEFG 1400
OY 1250 CKIIDGVRISDGLADLRKSLTIIPQEPVLFSGTVSRNLDPPFYQYTDQIWDALERTHM 1309
DB 1401 RIITIDIGIAKPLHLTRSLRSLTIIDPVLFSGTIRNLDPEKCSSTLWEALIAQLK 1460
OY 1310 ECTAQLPLKLESEVMEGDNFSGVEROLLCIARALLHCKIILLIDEATAMDDETDLIO 1369
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 9, 2001, 17:46:47 ; Search time 108.06 Seconds
(without alignments)
1759.414 Million cell updates/sec

Title: US-09-528-031-2
Perfect score: 7308
Sequence: 1 MMDIDGKRYIIPSPCYRSV.....DSRRFYAMFAAENKVAVKG 1437

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 2371.5 | 32.5 | 1427 | 5 062170 | 062170 caenorhabdi |
| 2 | 2082 | 28.5 | 1515 | 4 09UQ99 | 09UQ99 homo sapien |
| 3 | 2066 | 28.3 | 1545 | 4 09UMS2 | 09UMS2 homo sapien |
| 4 | 2060.5 | 28.2 | 1543 | 11 09J1H8 | 09J1H8 mus musculu |
| 5 | 2055.5 | 28.1 | 1528 | 11 035379 | 035379 mus musculu |
| 6 | 2038 | 27.9 | 1502 | 6 09GK09 | 09GK09 canis famli |
| 7 | 2022.5 | 27.7 | 1547 | 5 09I7N0 | 09I7N0 drosophila |
| 8 | 2022.5 | 27.7 | 1591 | 5 09VK56 | 09VK56 drosophila |
| 9 | 2013.5 | 27.6 | 1622 | 10 024635 | 024635 arabidopsis |
| 10 | 2013 | 27.5 | 1528 | 5 09N2N3 | 09N2N3 c.c. elegan |
| 11 | 2005 | 27.4 | 1540 | 5 094136 | 094136 caenorhabdi |
| 12 | 1993 | 27.3 | 1623 | 10 048907 | 048907 arabidopsis |
| 13 | 1992 | 27.3 | 1623 | 10 064590 | 064590 arabidopsis |
| 14 | 1988.5 | 27.2 | 1525 | 5 020943 | 020943 caenorhabdi |
| 15 | 1984.5 | 27.2 | 1494 | 5 020943 | 020943 caenorhabdi |
| 16 | 1972 | 27.0 | 1623 | 10 022449 | 022449 arabidopsis |
| 17 | 1959 | 26.8 | 1368 | 10 09S9R0 | 09S9R0 arabidopsis |
| 18 | 1953.5 | 26.7 | 1622 | 10 048908 | 048908 arabidopsis |
| 19 | 1897 | 26.0 | 1573 | 5 093552 | 093552 caenorhabdi |

| | | | | | |
|----|--------|------|------|-----------|--------------------|
| 20 | 1882 | 25.8 | 1327 | 5 09VLN6 | 09VLN6 drosophila |
| 21 | 1882 | 25.8 | 1539 | 10 09LZJ5 | 09LZJ5 arabidopsis |
| 22 | 1865 | 25.5 | 1459 | 4 09UQ97 | 09UQ97 homo sapien |
| 23 | 1858.5 | 25.4 | 1456 | 4 09UQAO | 09UQAO homo sapien |
| 24 | 1857 | 25.4 | 2580 | 5 09VJ21 | 09VJ21 drosophila |
| 25 | 1842 | 25.2 | 1465 | 3 09P5N0 | 09P5N0 schizosacch |
| 26 | 1830.5 | 25.0 | 1514 | 10 09LKE4 | 09LKE4 arabidopsis |
| 27 | 1828 | 25.0 | 1514 | 10 065619 | 065619 arabidopsis |
| 28 | 1819.5 | 24.9 | 1516 | 10 024525 | 024525 arabidopsis |
| 29 | 1818 | 24.9 | 1545 | 11 09JU67 | 09JU67 ratius norv |
| 30 | 1815.5 | 24.8 | 1306 | 10 09LU34 | 09LU34 arabidopsis |
| 31 | 1811.5 | 24.8 | 1348 | 5 09VDR3 | 09VDR3 drosophila |
| 32 | 1811 | 24.8 | 1515 | 10 024510 | 024510 arabidopsis |
| 33 | 1809 | 24.8 | 1513 | 4 09NXY3 | 09NXY3 homo sapien |
| 34 | 1787.5 | 24.5 | 1307 | 5 09VGM1 | 09VGM1 drosophila |
| 35 | 1784 | 24.4 | 1546 | 11 09U170 | 09U170 mus musculu |
| 36 | 1782.5 | 24.4 | 1545 | 11 063563 | 063563 ratius norv |
| 37 | 1781.5 | 24.4 | 1549 | 4 060706 | 060706 homo sapien |
| 38 | 1772.5 | 24.3 | 1490 | 10 09MLC7 | 09MLC7 arabidopsis |
| 39 | 1768.5 | 24.2 | 1493 | 10 09LKE2 | 09LKE2 arabidopsis |
| 40 | 1767.5 | 24.2 | 1346 | 5 09VANS | 09VANS drosophila |
| 41 | 1767 | 24.2 | 1144 | 5 09XW61 | 09XW61 caenorhabdi |
| 42 | 1763.5 | 24.1 | 1549 | 6 082451 | 082451 oryctolagus |
| 43 | 1752.5 | 24.0 | 1582 | 11 070532 | 070532 ratius norv |
| 44 | 1747.5 | 23.9 | 1582 | 11 09EOT0 | 09EOT0 ratius norv |
| 45 | 1727 | 23.6 | 1344 | 5 09V572 | 09V572 drosophila |

ALIGNMENTS

```
RESULT 1
ID 062170 PRELIMINARY; PRT; 1427 AA.
AC 062170; 062241; 09U095;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE HYPOTHEICAL 158.7 KDA PROTEIN F14F4.3 IN CHROMOSOME X.
GN F14F4.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis J.M.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE FROM N.A.
RA Percy C.M.;
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
CC -I- ALTERNATIVE PRODUCTS: TWO FORMS; ISOFORM F14F4.3A AND ISOFORM
CC F14F4.3B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AL021446; CAB54225.1; -.
DR EMBL; 283227; CAB54225.1; JOINED.
DR EMBL; 283227; CAB54251.1; -.
DR EMBL; AL021446; CAB54251.1; JOINED.
DR EMBL; AL021446; CAB54226.1; -.
DR EMBL; 283227; CAB54226.1; JOINED.
DR EMBL; 283227; CAB54250.1; -.
DR EMBL; AL021446; CAB54250.1; JOINED.
DR HSSP; P13569; INED.
DR InterPro; IPR001140; -.
DR InterPro; IPR001687; -.
DR InterPro; IPR003439; -.
DR Pfam; PF00005; ABC_tran; 2.
DR Pfam; PF00664; ABC_membrane; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN.2.
KW Hypothetical protein; Alternative splicing.
FT VARSPLIC 643 669 MISSING (IN ISOFORM F14F4.3A).
SQ SEQUENCE 1427 AA; 158685 MW; 81DA0FBC90961FAE CRC64;
```

| Query Match | 32.5% | Score 2371.5 | DB: 5 | Length 1427 |
|-----------------------|------------------|---|------------|-------------|
| Best Local Similarity | 36.3% | Pred. No. 1,2e-152 | | |
| Matches 507 | Conservative 289 | Mismatches 497 | Indels 103 | Gaps 20 |
| QY | 51 | ETTAAREGISLSDASHMSLOLRLLDEHPHGKKNHNGSALPKPIETCKHOH---- | PVDNAGL | 106 |
| Db | 26 | ETSDRAKRYADDTNKKT-----IGRSAAYQNLIPLTTERKNNNGSGSRIDDAGL | 75 | |
| QY | 107 | FSCMTFSLSSLARVAHKKGELSMEDWVSLSKHSSDYNCRRLEFLMOBELNEVGPDAAS | 166 | |
| Db | 76 | FSFVYVSWVFPEYLXOA-VRGKIDRNQWVGCSEYDSCGLMRLLEWLMEDE-KKANAKSPS | 133 | |
| QY | 167 | LBRVWVICRRLLIISTYICMLMITOLAGSGPFWMKHLELYQATVESNQ-----YSL | 220 | |
| Db | 134 | LKRYIRISTRLMFSCAVFFECFLIFGIIGPCFTRRLIARFENERPDSQIIVSYGIA | 193 | |
| QY | 221 | LVGLLLEIYRSWSLALTMALNPTGVRLNGALITMAFKLIKLIKNTKEKSLGELINIC | 280 | |
| Db | 194 | LVAISVVEPARVLISYGAMVAVSYGTGRVNGAVLALLXKNVLSKDLCKGKESDVINIF | 253 | |
| QY | 281 | SNDGQRMFEBAAVGSLAGCPVVALIGMIVNYIIGPFGELGSANFIFLFPYMPAFASRLT | 340 | |
| Db | 254 | ANDGRLPDATYFADLVYGVPLVGLVGGIYGLMTVIGRMSLGLIWFEPEDVYIQFGLSM | 313 | |
| QY | 341 | AYFRKCAVADDEROKMNEVLTLYKFIKMAVMAVFAOSVOKIREEBERILIEKAGFOS | 400 | |
| Db | 314 | VACRLALVTKETKRSMAAEIKYIRYKKNMDEIFSAKIDQFKREKEVOYIRKSGYAS | 373 | |
| QY | 401 | ITVGAPIVNVVIAVYTESVHNTLGFDLTAAQAFVTVTVEFNSMFAKVTPEVSKLSFA | 460 | |
| Db | 374 | LAIAGCPVPVVAALITFVGVVLALNDLILASPAFSAITVYFMLEGIRIMIPGSRYLEA | 433 | |
| QY | 461 | SVAVDRFSLFMEBV--HMKNKPAASHIKEMKNATLAMSSSHSSIONSKLPIPKMK | 518 | |
| Db | 434 | VYAMRIDEIYLLIEDYAPRYTN--AEDVYIDCGAATTYO-----PK---- | 474 | |
| QY | 519 | DKRASRGKKEKYROLQRTHEQAVLAEOKGHLLDSDRPSPEEBGKHITHGLRLOTL | 578 | |
| Db | 475 | --AAKAPVDETK--EPTENEIVYETPVF-----TC | 501 | |
| QY | 579 | HSIDLEIDGKLVIGCGSVGSKTSLISALIGOMLLBSIAIS-GTPAYVAQAQMIINA | 637 | |
| Db | 502 | SFDKLSIRGEHIAVIGAVCGGSAILKIASHMTPTTDALSDVRSQTYVYVOKMIFNG | 561 | |
| QY | 638 | TLRDNIITFGEKEDERYAVSLNSCCLRPDLALIPSSDLIEIGERANLSGGORISLAR | 697 | |
| Db | 562 | TYQDNILITGDKKNSRERYKKAIVGCCOLTELTLTSLGDRIVEGERGATISGGAKAVALAR | 621 | |
| QY | 698 | ALYSRSIYIILDDPLSALDAVYGNHIFNSAIRKHLKSKTVLFVTHOLOLYLVDCDEVIEMK | 757 | |
| Db | 622 | AVFQIKNLYIFEDDIEASLDKVAANKIHEBIILOKLKKALAMVTNNMELIHFDRYLEFE | 681 | |
| QY | 758 | BGCITERGHEEIMLNDGYATIFNNLLIGETPVEINSKTEPSSQOKKSGPKPTGSI | 817 | |
| Db | 662 | GSNIYADGNHDLILYKNDAYKTFVD-----ACETIYATSGATSICGPGAPQAPAL | 731 | |
| QY | 818 | KKE-----KAVPEEGOLVLEE-KGGSYVSWSYGYVYIQAAGPLAFLVIALFMLNV | 870 | |
| Db | 732 | DAEILRNSESDIKGADKALISDEEDMGNSTIMARIYKQYIIHAAGWPITWCIVGIFIVN | 791 | |
| QY | 871 | GSLASTMWLSWIMOG--SGNTTVRGNF--TSVDSMKDNPMQVYASITALSMAVMILL | 928 | |
| Db | 792 | VSNIFSTYMLSKKGGHDETTIYNGTEFEEMKTSIADSPYTGFAAYVALVALVLTIS | 851 | |
| QY | 929 | KAIRGVVPEVKGTLARASSRLHDELFRILRSPPMKFPTDPTPTGRILNFRSKMDDEVDRILPE | 988 | |
| Db | 852 | GLFKACVPEVKSILFATRIHRDHMPQAVIHGATSFDSPTPTGRILNFRSKMDDEIVKLPE | 911 | |
| QY | 989 | QAEMLQNVILVFCVG---MAGVFPWFVAVGVPLVILFSLVHIVSRVILELRKLDNI | 1045 | |
| Db | 912 | TAELVNLNMIT---CLGELVLTIVTFEPEFLFALIFLVEFVVEVFCFRGIRLNKSEHI | 968 | |

| | | | |
|--------|---|--|------|
| QY | 1046 | TQSPFLSHITSSIOGLATITAYKKGEFLHRYVELLDNDQAEFFLFTQCMRLAVRLDI | 1105 |
| QY | 1046 | TSQSPFLSHITSSIOGLATITAYKKGEFLHRYVELLDNDQAEFFLFTQCMRLAVRLDI | 1105 |
| Db | 969 | SRSPLEHYHVSASLEGITITTFPOOSNRFFELVLRKHLDCNSGALFMQSAMRLAVMLDL | 1028 |
| QY | 1106 | SIALITTTGLMIVLMHCOIPAYAGLAISAVOGLTGPOFTRLASETEARFTSEVERINH | 1165 |
| Db | 1029 | VVWMTAIVALLITVWLITCTVSPADAGMALAFAYVMOSSIIPOFNARTQTELEAKNTSEVERSY | 1088 |
| QY | 1166 | YIKTLSEAPARIKNKAP-SPMDPOGEVTFEENAEKRYEENLPLVLYKKVFTIKREKIG | 1224 |
| Db | 1089 | YADNIPEDGEMNTHQGLDISSWPNANQINFSEVNLRYKRSHPLALNDITFEIKGGEKYG | 1148 |
| QY | 1225 | IVGRTGSKSSSLGMAFLRVLVELSGCCIKIDGVASIDIGLADLSKISITPOEPVLFSGTV | 1268 |
| Db | 1149 | IIGRTGSGKSSSLAMILIRFLPVTNGITVYIDGVIRFVGVLKLRIGISAIAOPDSLSEGTV | 1208 |
| QY | 1285 | RSNIDPFNOGYEDQIMPALERTHMKKECIAQPLKLESEWENGDNFSVGEROLLCTARAL | 1344 |
| Db | 1209 | RPNIDPSPLEYSDSNIMWEALRCHKITLVOSLDKKLADVSHGCGNFSVGEROLFCLARAL | 1266 |
| QY | 1345 | LRRCHKILIDEGATTAAMOTETDILLIOETIRFAPDCITMLIAHRLITVLSGDSINVLAOGQ | 1404 |
| Db | 1269 | LMKSRVITLDEATATSVAGNDKILQEVYIKVFEADAVIIIAHRLDVRMNDIMHLKNKG | 1328 |
| QY | 1405 | VVEFDTPSVLLSNDSS 1420 | |
| Db | 1329 | LINEFTTPOEMFKDDMS 1344 | |
| RESULT | 2 | | |
| Q90099 | Q90099 | PRELIMINARY; PRT; 1515 AA. | |
| AC | Q90099; | | |
| DT | 01-MAY-2000 (TREMblrel. 13, Created) | | |
| DT | 01-MAY-2000 (TREMblrel. 13, Last sequence update) | | |
| DT | 01-MAR-2001 (TREMblrel. 16, Last annotation update) | | |
| DE | MULTIDRUG RESISTANCE PROTEIN (FRAGMENT). | | |
| GN | MRP. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Euthera; Primates; Catarrhini; Homnidae; Hom. | | |
| OX | NCBI_TaxID=9606; | | |
| RP | [1] | | |
| RNA | SEQUENCE FROM N.A. | | |
| RA | MEDLINE=98008927; PubMed=9344662; | | |
| RT | Grant C.E., Kurz E.U., Cole S.P.C., Deley R.G.; | | |
| RT | "Analysis of the intron-exon organization of the human multidrug- | | |
| RL | resistance protein gene (MRP) and alternative splicing of its mRNA."; | | |
| CC | Genomics 45:368-378(1997). | | |
| CC | -1 SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC | | |
| CC | TRANSPORTERS). | | |
| DR | EMBL; AF022853; AAB83979.1; -. | | |
| DR | EMBL; AF022824; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022825; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022826; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022827; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022828; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022829; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022830; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022831; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022832; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022833; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022834; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022835; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022836; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022837; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022838; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022839; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022840; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022841; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022842; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022843; AAB83979.1; JOINED. | | |
| DR | EMBL; AF022844; AAB83979.1; JOINED. | | |

| | | | | |
|----|--------------|------------|--------------------------------|--------------------------|
| DR | EMBL: | AF022845: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022846: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022847: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022848: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022850: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022851: | AA883979.1: | JOINED. |
| DR | EMBL: | AF022852: | AA883979.1: | JOINED. |
| DR | HSSP: | P13569: | INBD. | |
| DR | InterPro: | IPR000719: | - | |
| DR | InterPro: | IPR001140: | - | |
| DR | InterPro: | IPR001687: | - | |
| DR | InterPro: | IPR003439: | - | |
| DR | InterPro: | IPR003593: | - | |
| DR | Pfam: | PF00005: | ABC_tran; 2. | |
| DR | Pfam: | PF00664: | ABC_membrane; 2. | |
| DR | PROSITE: | PS00211: | ABC_TRANSPORTER; 2. | |
| DR | PROSITE: | PS00107: | PROTEIN_KINASE_ATP; UNKNOWN_1. | |
| DR | SMART: | SM00382: | AAA; 1. | |
| DR | ATP-binding: | Transport. | | |
| KW | NON_PER | 1 | | |
| FT | SEQUENCE | 1515 AA; | 169852 MW; | DC685592817C439FE CRC64; |

DR EMBL: AJ132287; CAB45309.1; JOINED.
DR EMBL: AJ245625; CAB45309.1; JOINED.
DR EMBL: AJ132288; CAB45309.1; JOINED.
DR EMBL: AJ132289; CAB45309.1; JOINED.
DR EMBL: AJ132290; CAB45309.1; JOINED.
DR EMBL: AJ132291; CAB45309.1; JOINED.
DR EMBL: AJ132292; CAB45309.1; JOINED.
DR EMBL: AJ132293; CAB45309.1; JOINED.
DR EMBL: AJ132294; CAB45309.1; JOINED.
DR EMBL: AJ132295; CAB45309.1; JOINED.
DR EMBL: AJ132296; CAB45309.1; JOINED.
DR EMBL: AJ132297; CAB45309.1; JOINED.
DR EMBL: AJ132298; CAB45309.1; JOINED.
DR EMBL: AJ132299; CAB45309.1; JOINED.
DR EMBL: AJ132300; CAB45309.1; JOINED.
DR EMBL: AJ132301; CAB45309.1; JOINED.
DR EMBL: AJ132302; CAB45309.1; JOINED.
DR EMBL: AJ132303; CAB45309.1; JOINED.
DR EMBL: AJ132304; CAB45309.1; JOINED.
DR EMBL: AJ132305; CAB45309.1; JOINED.
DR EMBL: AJ132306; CAB45309.1; JOINED.
DR EMBL: AJ132307; CAB45309.1; JOINED.
DR EMBL: AJ132308; CAB45309.1; JOINED.
DR EMBL: AJ132309; CAB45309.1; JOINED.
DR EMBL: AJ132310; CAB45309.1; JOINED.
DR EMBL: AJ132311; CAB45309.1; JOINED.
DR EMBL: AJ132312; CAB45309.1; JOINED.
DR EMBL: AJ132313; CAB45309.1; JOINED.
DR EMBL: AJ132314; CAB45309.1; JOINED.
DR HSSP: P13569; 1MBD.
DR InterPro: IPR001140; -.
DR InterPro: IPR001687; -.
DR InterPro: IPR003439; -.
DR InterPro: IPR003593; -.
DR Pfam: PF000005; ABC_tran; 2.
DR Pfam: PF00664; ABC_membrane; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR SMART: SM00382; AAA; 1.
DR SEQUENCE 1545 AA; 174148 MW; 1A036686FB1628FC CRC64;
SQ
Query Match 28.38; Score 2066; DB 4; Length 1545;
Best Local Similarity 34.58; Pred. No. 9.5e-12;
Matches 501; Conservative 244; Mismatches 491; Indels 218; Gaps 27;
99 HPVNAGLFCSCMTFSWLSLARVAHKKGELSMEDYWSLSKRESSDVNCRLERLMOEELN 158
193 NPSSIASPLSSITTSWDSIILKGYKR-PLTLEDVWEVDEMKRTTLVSKFETIHKRELO 251
159 EV-----GPDASLRV-----VN-- 172
252 KARALDRQROEKSSQONSQARLPGLNNQOSQDALVLEDEYKKKKSGTKKDYPKSYM 311
173 --TICRRLLI--SIVCLMTQLAGFSPAPMVKHLEYTATESNDLOYSLILVIGLLIT 228
312 KALKTFEVMVLKLSFLFLKLVNDIFTFVSPQ-LIKLILSFASDRDTYLMIGYLCALIFTA 370
229 EIVSWSLATLWALNRYRGVLRGAILTMARKKILIKLNI--KEKSGELINIGNSDOR 286
371 ALISFCIQCFQLOCFKLGVRTRAIMASVYKKALLTSLNLRKRETYGVEYVNLMSVDNOK 430
287 MFEAAVSLAGPVPVAVLGMIVYVILIGPTGLSGLAVILFYPAMFASRLTAFFRRK 346
431 LMDVTNFMHMLMSSVQLVLSIFFLMRGLSPVLAGVGMVLVPIINALISTKSTIQVK 490
347 CVAATDERVQKMEVLYIKFKIKMYAMVAKASQSVOKITREERIRLLERAGYFQSTIVGA 406
491 NMKNKDKRLKIMNELLSGIRLIKTFYAWEPSPRDQYONLRKKELKLLAFSOLQCVYIEVF 550
407 PIVVIVIASVYTFYSYHMTIGFD--LTAAGATVVVYVFNSTALAKTTPPSVVSLSASAV 464

551 QLPFVLVSVYTFYSYVLYVDSNNILDAQKFTSTILFNILRPPLSMLPMISSLQASVST 610
465 DREPSLIMEVYHMTIKKPPSPH-----IKEMKALNLANDSSHSSTIONSPKILPKKKKD 519
611 ERLEKYLGGDL-----DTSAIRHDCNFDKAMQFSEAFSTW----- 646
520 KRASRGKKEKVRQLQRTHEQAVLAEOKGHLLDSDERPSPEEBEGKHILHGLRLQLH 579
647 -----EHDSS-----EATVR 655
580 SIDLEIEGKLVIGCGSVGSKTSLISAILGQMTLLEBSIAISGTFPVAQAMTANL 639
656 DVNLDIMAGQLVAVIGPVGSKSSLSIAMLGEMENVGHITIKGTAYVPOQSWIQGTI 715
640 RDNILFKEVDEERYNSVLANCCRLPDLATIPSSDLIEIGRGANLISGGQORISLAL 699
716 KDNILFTEFEKRYQOVLFCALLPDLLEMIPGGLDLIEIGKINLSGGOKRISLAKAT 775
700 YSDRSYIIDLPLSALDAVGNHIFNSAIRKH--LKSQTVLFYTHQLOLYVDCDEVIFMK 757
776 YONDIYILDDPLSAVDVAHVGNHIFNKVYLGPNGLLKGTRLTYHSMHFLPQVDEIYVLG 835
758 EGCITRGTHIELNLNGDYATIFNNLLGETPVE----- 793
836 NGTIVEKGSYALLAKKGEFAKNLKTFLRHTGPBEAATVHDGSEEDDYGILSSVEIP 895
794 -----INSKRTS-----GSOKKSQDXGPKTGSI-----KKEKAYKPEGQLOVQ 832
896 EDAASITWRRENSFRRLTSRSSRNGRLKSLRNSLKRNNVSLKEDEELVKGQ--KLTK 953
833 LEEKGQGSVPWSYGVYVIOAGGFLAFLVIALFMLNNGSTAF--STWMLSYWIKQSGN 890
954 KEPIETGKVFESIYLEYQALGIFSIFPIILA-FVMN--SVAFIGSNMLMSAM----- 1003
891 TTVTRGNETSVSDSKMDPNHMOYASISYALSMAVLLIKAIRGV-VFYK-----GITRA 943
1004 -----TSDSKIFN-SDYPASQDMRVGVYGAALGQIFVFIAHRWSAFGFVHA 1052
944 SSRHLDELFRILNSPMKFEPTTPGRLNFRSKMDDEVDRLPQAMFIONVILVFC 1003
1053 SNILHKKLLNNILNLMKMFETTPGRIVNRFPAGDISVDDTLTPOSLSRWTCFLGIIST 1112
1004 VGMIAVPEWFLVAVGPIVLIIFSVLYSVRLIRELKRDLNITOSPFLSHITSSIOGIAT 1063
1113 LVMICMARFVFTIIVIPGIIIVSYQMFYVSGTRQLRDLSDVTRSPISHSSEVSGLPV 1172
1064 IHAYNKGGEFLHRYQELLDDNQAPFLEFTCAMRLAVRLDLISALTITTTGLMTVLMHGQ 1123
1173 IRAPFHQORFLKHNNEVRIDTQKCVFSWITSNRMWLAIKLELVGNLTVFFSALMVIYRDT 1232
1124 IPPAVAGLAISYAVQLTGLFQFTVYLAETEARLTSVERINHIYIKTSLSLDEAPARIKNKAP 1183
1233 LSGDTVGFVLSNALNITVTLNWLVMYMTSEIFENIVAVRIETRYK-VENEDP-WYTDKRP 1290
1184 SPDPPOGEVTFENAEVRENLPLVLRKVSFTIKPKKIGIVGTGSGKSLGALFRL 1243
1291 PDPWPSKGIQFNNTYVRRRLDLVLGRTICDIOGSMKIGVGTGSGKSLTNCLEFRI 1350
1244 VELSGGCIKIDGVRISDGLADLRSLKSLIIPOEVLVFGTVRSNLDPPNOYTEDQIWDAL 1303
1351 LEAAGGQIITIDGVDLASGLDLRKLITIPQDPLIFGSLRMLNDPPNNYSDEEIMKAL 1410
1304 ERTNKECIAOLPLKLESEVWENGDNESVGEROLCIRALLRCKIILIDETAAAMTE 1363
1411 ELAHLKSFVASIQGLSLHSGTEAGGNLSIGRQLCLICLRALLRKRIILVLEDEATAVDLE 1470
1364 TDLLIQTETREAFACVMTLTAHRLHTVLGSDRIWVLANOGOVVEDEPVSLLSNDSS-RF 1422
1471 TDNLIQTITQNEFACTVITTAHRLHTIMDSKVWVLDNGKIIIEGSPDELLQIPGPYF 1530
1423 YAMFAAENKAVK 1436
1531 MAKEGIENVNSTK 1544

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durdin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B., Kodish C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
The genome sequence of *Drosophila melanogaster*.
RL Science 287:2185-2195(2000).
CC - SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
TRANSPORTERS).
DR EMBL: A003637; AAC22430.1; -.
DR InterPro: IPR000719; -.
DR InterPro: IPR001140; -.
DR InterPro: IPR001687; -.
DR InterPro: IPR003439; -.
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00664; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR KAP-Binding; Transport.
KW SEQUENCE 1547 AA; 173190 MW; 492BD9AF05B80681 CRC64;
Query Match 27.78; Score 2022.5; DB 5; Length 1547;
Best Local Similarity 33.08; Pred. No. 8; Be-129;
Matches 497; Conservative 256; Mismatches 455; Indels 299; Gaps 34;
80 KYHGLSALPRTCTKHOHPVDNAGLFSCFTFSLSLARVAHKKGELSMEDWWSLSKH 139
172 KYQNG-----ENRDELSSASFLSRTTYQMPDKMAKGR-RNPLEKDKMDLDRPQ 219
140 ESSDVCNCRLELMQOELENE-----VGPD-----ASLRR 169
220 DSCSEVPPIFAHNNQNVKRYKKARKEPKAOPNSNGNVTFENPHEGKNGKMKMASIMP 279
170 VVMFECFTRILTSIVCLMIOQLAGEGPAFVYKLLLEYTQATESNLDQSYLLVGLILTE 229
280 PIYSEFGVFEFGALMKLEFDLTTFAPQ-VLSLISFVEQDEPEPK-----GLITAV 333
230 IVRSMSLALPRTCTKHOHPVDNAGLFSCFTFSLSLARVAHKKGELSMEDWWSLSKH 281
334 LFLYLAAGAFILQGYFHRMFIYGLIRITLALINAYRKALISNSTKKESTVEGLVNLMA 393
282 NDGGRMFEEAAGVSLAGPVAALIGMIVNYILGPTGELGSAVFLEFPAM-MFASRLT 340
394 VDAQREFELTYLMMIMSAFLQIGLALYFLMOQLGSPVLAGLAVIILIPVNGVIASRIK 453
341 AYFRKKAATFDEVOMNEVLTYIKFKIMYAWYKAFSOSQKIREERRRLLEKAGYFOS 400
454 TYQIRQ-MKYDERVKLMEVLSGIVKLTAMEPSEKQVLDIRKXIATLSTAYLNA 512
401 IT-----VGADIVVIA----- 413
513 GTSLMSCAPFLSVLYFAFVYVLTSEANQLSEKIVIVISIALFDLMKLPRTITLPMISVDA 572

QY 414 --SVYFSVAMTLCFD--LTAQAFTVYVYVNSMTFALKYTPSVYSLSPASVADRFKS 469
D 573 EVSLYFATYVLLDENNVDAATPFTFSLISFNILRFLPLMILITNLVQVSVNINK 632
QY 470 LFLMEY--HMIRKPAFSPKIKEMKATLMDSSHSIIONSPLFPMKKDKRASRKK 527
D 633 FLNSELDPVSLHDSKRP-PMSENGEFSW----- 663
QY 528 EKVROLORTEHOAVLAQKGLHLLDSDRSPSEEEKHILHGLRLQRTLSIDLEIQE 587
D 664 -----GDEI-----TLRNIIEVK 678
QY 588 GKLVIGSGVSGKTSLSALIGOMTLBESIALSGFAVVAQOAWTLNTRDNLIFGK 647
D 679 GSTVALGTGSGSGSVQAFLEMEKLAQVNTGVLAVPQDAWQONATVRNLLIEGQ 738
QY 648 EYDERYNSVLNSCCLPDLALPSSDLTEIGERGAALSGGORISLARALYSDRSITY 707
D 739 TYDKRRYKVIKCALADADILSAGDLTEIGEGTILNSGQOKORISLARALYSDADLYL 798
QY 708 LDDPLSALDAHVNHNISAL--RKHLKSTVLEVTQLOLYVDCDEVIFMKEGCITERG 765
D 799 LDDPLSALDAHVGHIPEEYIGPKILARKSRVLYTGVTFLPQVDISYIKMEIISG 858
QY 766 THEELMANGDYAL-IRNNLILGETPPEVEINSK-----ETGSGQKK----- 806
D 859 TFDPLVKNKGAFAFDIYIHLQEGNEEELNQRKQISSADVPPELLGTVEKAIKARTE 918
QY 807 -----SODGPKTGSIKKEKAVKPEEGOLVQLEBK 837
D 919 SLDSISVTSADSLMGGGSLRRRTKQDSHDSVASAASLKKQEV--EGKLETERKQ 975
QY 838 QGVPMVSVGYVYIQAAGPLAFVIMLMLNGSTAF--STWMLSYIKQSGNTTFT 894
D 976 TGGEVAFVYKHYIKSVG---IFLSVATLVLFVFOAFQJSGNSMLTQW----- 1020
QY 895 RGNETSVDSKNDPMMQYATYAL-----SNAVMILKAIKGVYVVKETLR 942
D 1021 -ANDQNVAN--DTGLRDMVLGYGAFGQVATNFSSLAISL-----GCLK 1064
QY 943 ASSRLHDELFRILRSPMKPFDDTPTGRLNRFSDKODEVDVRLPFOAEMFIQWILVFF 1002
D 1065 CSOLHQTLLYNNLRWPELEFDTTPLGRIVAKFSKQIDITDNLVFNIRVYIGAYVLA 1124
QY 1003 CVGMIAGVFWFLVAVGPVLYLFSVLHIVSVLI--RELKRDNIQSPFLSHITSIQ 1059
D 1125 TIVVISLSTPIFLAVIVPIAFLY--YFAQRFVATSRQLMRLESVRSRPIYSHFSETVT 1181
QY 1060 GLATTHAVNKGQERLHRYQELLDONQAPFLFTCAMRLAVRDLISALITTTGLMIVL 1119
D 1182 GASTIRAYNGVDRIESDAKVDKNQCKYPSVIANNRLAIRLMEVGNLILFASLAVL 1241
QY 1120 MHGQIPPAYAGIAISYVQLTGLFQFVYRLASETEARFTSVIERINHYIKTSLPARIK 1179
D 1242 -GGQTNGLVGLSVYALQVYQTLNMLVYRMSDDIETINIVSERIKKEGET-KQAPPELE 1299
QY 1180 NKADSP-DWPOEGEVTEBNAMRYRENPLVLKAVYFTIKPKKEIGYVGTSGSKSLGM 1238
D 1300 QDKRKPMPQEGVEHVEONFQVYRREGIDLVRGVSINQEGEYVGIYGRGAKSSITL 1359
QY 1239 ALFRVLVSGGCIKIDGVRISDIGLADRSKLSIIPQEVFSGTGVSNLDPEQNYEDQ 1298
D 1360 ALFRILAEAGRISIDGVASLMSGLHMRSLRTIIPQPVLFSSLRNLNDPEFIKTDE 1419
QY 1299 IWDALERTHMKECIAOLPLKLESEVMGDNFVSERQOLIAALRLHCKIILIDETA 1358
D 1420 IWKALDELHLSKFSYKSLAAGINHLAEGENLSVGQRLVQCLARALLKRPVVLDEATA 1479
QY 1359 AMDTETDLLIQETIREAFADCTMTLTAHRLHTVLSDRIMVLAQGVVEEDPVSLLSND 1418
D 1480 AVDETETDLLIQ-----FKECTVLTIAHRLNTIIDSQVIVLDGQIIEFASPELLDNP 1533

| RESULT | 8 | | |
|--------|---|------|----------|
| 09VK56 | PRELIMINARY; | PRT; | 1591 AA. |
| AC | 09VK56. | | |
| DT | 01-MAY-2000 (TREMBLrel_13, Created) | | |
| DT | 01-MAR-2001 (TREMBLrel_16, Last sequence update) | | |
| DT | 01-MAR-2001 (TREMBLrel_16, Last annotation update) | | |
| DE | CG6214 PROTEIN. | | |
| GN | CG6214. | | |
| OS | Drosophila melanogaster (Fruit fly). | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| XX | NCBI_TaxId=7227; | | |
| XX | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=BERKELEY; | | |
| RC | MEDLINE=20136006; PubMed=10731132; | | |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D., | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abbil R.M., Asgari A., An H.-D., Andrews-Flannkoch C., Baldwin D., | | |
| RA | Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P., | | |
| RA | Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablos B., Delcher S., Deng Z., Mays A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., | | |
| RA | Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., | | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., | | |
| RA | Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., | | |
| RA | Jatall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z., | | |
| RA | Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., | | |
| RA | Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., | | |
| RA | Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A., | | |
| RA | Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., | | |
| RA | Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M., | | |
| RA | Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G., | | |
| RA | Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., | | |
| RA | Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., | | |
| RA | Spieler E., Spreading A.C., Stapleton M., Strong R., Sun E., | | |
| RA | Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., | | |
| RA | Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., | | |
| RA | Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., | | |
| RA | Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., | | |
| RA | Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Zhu X., | | |
| RA | Gibbs R.A., Myers E.W., Zhang G.M., Venter J.C.; | | |
| RL | "The genome sequence of <i>Drosophila melanogaster</i> ."; | | |
| RT | Science 287:2185-2195(2000). | | |
| CC | -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC | | |
| CC | TRANSPORTERS). | | |
| CC | EMBL; AE003637; AAF53223.2; -; | | |
| DR | HSSP; P13569; INDB. | | |
| DR | Playbase; FBgn0032456; CG6214. | | |
| DR | InterPro; IPR000719; -; | | |
| DR | InterPro; IPR001140; -; | | |
| DR | InterPro; IPR001687; -; | | |
| DR | InterPro; IPR003439; -; | | |
| DR | Pfam; PF00005; ABC_tran; 4. | | |
| DR | Pfam; PF00664; ABC_membrane; 3. | | |
| DR | PROSITE; PS00211; ABC_TRANSORTER; 2. | | |

| DR | PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.. | 27.7%; Score 2022.5; DB 5; Length 1591; |
|----|---|---|
| KM | ATP-binding; Transport. | Best Local Similarity 33.0%; Pred. No. 9.2e-129; |
| SQ | SEQUENCE 1591 AA; 1/8384 MW; 735AB6DA895BA859 CRC64; | Matches 497; Conservative 256; Mismatches 455; Indels 299; Gaps 34. |
| QY | 80 KYNHGLSALKPIRTTCKHQHPVDNAGLFSQCMFSLWSLSLARAHHKGLSMDVSLSKH 139 | |
| DB | 216 KVGQG-----ENRIPELSSFLSRLITYQMPDKNALGY-RNPLEEKDMLDRQ 263 | |
| QY | 140 ESSDVNCRRLKQLEELNE-----VGPD-----ASLR 169 | |
| DB | 264 DSCSEMPRTFAHHMNQNNKKNKRAVEPKAQFSGNVTFFENPHGEKNGRRKGMAISMP 323 | |
| QY | 170 VVWICFRPLILSIYCIAMTLAGFSGPAFWKHLLEYQATEESNLQYSLLVGLLRE 229 | |
| DB | 324 PIYKSPGVFLGLAMKLEFTDPLTLTAQPO-VLSLIISFEADADEPK-----GLTAV 377 | |
| QY | 230 IVRSMSLATLWALN-----YRTGYRLGCAILLMAFKKILKLKNTI--KEKSGELINICS 281 | |
| DB | 378 LLFVLAQAQFTLGLQYFHRMFIVGLRIRALLNALYRKALRISNPKKSTYGEIYNMA 437 | |
| QY | 282 NDQGMFEAAVAGSLLAGRPVAILGMITYNIIIGPTGLSAVFLFLPAM-MRASRLT 340 | |
| DB | 438 VDAQFMELTTLYNIMWSPALDIGALYFLMOQLGSPVLAGLAVMILLPVNGVIAIRIK 497 | |
| QY | 341 AYFRKRCVATDERPOKMEVLTLYIKFIKVMVAKFASQVOKIIEERERILEKAGYFQS 400 | |
| DB | 498 TYQIQH-KMYRDERKALNEVSLGIKLYLAMEGSEFEQVLDIRKELATLRSTRAYLNA 556 | |
| QY | 401 IT-----VGAPIVTVIA----- 413 | |
| DB | 557 GSFIMSCAPFLVSLVTFATYVLTSEANQLSEKVLISIALDLPMKPLTLIIPMLSVDA 616 | |
| QY | 414 --SVYTFSVHMLGSD--LTAQAFTVTVYFNSMFTALKVTPFSVKSLSAASVANDREFS 469 | |
| DB | 617 EYSLVTFATYVLLIDENNVDATKFTVSLSLEFILRLPLMLPLTLNLVOTOVSVARINIK 676 | |
| QY | 470 LELMEEV--HMIKRNPASPHIKIEKKNAVLADSSHSISSIONSPKLTPKMKKDRASRGK 527 | |
| DB | 677 FLNSELDPNSVLHSSKSH-PMSENGEFMS----- 707 | |
| QY | 528 EKVROLQRTHOAVLAEOKGHLLDSDERPSPDEEKGHIHLGHLRLQTLHSIDLEIOE 587 | |
| DB | 708 -----GDEI-----TLRNINIEYVK 722 | |
| QY | 588 GKLVIIGSVSGSKTSLISALIGOMTLLEGSLAISGTFYVAQOAMINATLRDNLPEK 647 | |
| DB | 723 GSLVALVGVSQKSSVQVAFGEHEKRLAGVNVYGLKLVYPOQAMIONATVRDNLTFQO 782 | |
| QY | 648 EYDERRYSVLSNCCLRPLDILPSSDLTEIGERGANSIGGORISLARALYSRSRIY 707 | |
| DB | 763 TYDRKRYNKVIDACALRADRIDLSAGDLTEIGEEKGINSGGKQRIISLARAVYSQADLYL 842 | |
| QY | 708 LDDPLSALDAVYGNHINENSAT--RKHLSKTYLVYTHQLOLYLVDCDEVYFMKEGCTTEKG 765 | |
| DB | 843 LDDPLSAYDAVHAGHKIFEEVIGPKDILARKSSVLYLTHGVTLPOVDYSIVIMGEISSEG 902 | |
| QY | 766 THEELMLNNGDYAT-IFNNLLIGETPPVEINSK-----ETSGSQRK----- 806 | |
| DB | 903 TFDQLVKKNKGAFTIIGHLQEGNEEBEELNIKKQISSTADVPPELLGTIVEGAIKLART 962 | |
| QY | 807 -----SODGPKTGSIKKEKAVKPEEGQVOLAEEKG 837 | |
| DB | 963 SLSDISVTSADSLMGGGGSLRRRTKRODSHDSVASASLSKKOEY---EGKLIETEKQO 1019 | |
| QY | 838 QGSVWYVSYGVYIOAAGGGLAFVLVMALEFMLVGSSTAF---STWMLSTVIRKGSNGTYT 894 | |
| DB | 1020 TGGVEFAVYKHKIKSVG---IFLSVATLVLVNFQAFQISGNMLWTQV----- 1064 | |

QY 895 RGNETS DSKNDPNHMYASYAL-----SMAVMILKAIKGVVFKGTLR 942
 Db 1065 -ANDQYAN--DTGLDMYLGYGAFGEGQVATNFSSALISL-----GCLK 1108
 QY 943 ASSRLHDELFRRLIRSPKPFEDTPTGRIILNRSKDMDEVVRLPFOAEFIQVILVVF 1002
 Db 1109 CSQLLHGTLLLYNRMWMEFEDTTPGRIRYNRSKDIIDTIDNVPINRVIGAYVLA 1168
 QY 1003 CVMGIAGVFWFLVAVPPLVILSVLHIVSRVLI---RELKRDNTIQSPFLSHITSIQ 1059
 Db 1169 TIYVISTPTFLVAVPIAFVLY---YFAQRFVATSRQIMRLRESVSRSPYISHSETVT 1225
 QY 1060 GLAIIHYNNGQELHAYQELIDNDAPELFCAMRWLAIVRLDLSIALITTGIMIVL 1119
 Db 1226 GASTIRAYNNGDRFIEESDAKVDKNQCKPVSATIANRMALRLMVGNLILPASLPAVL 1285
 QY 1120 MHGQIPAYAGIAISYAVOFTGLFQFTVRLASETAEFTSVERINHITKTLSEAPARIK 1179
 Db 1286 -GGGTNGGLVGLSYVALQYLTINMLVYRMSDLETINIVSVERIKEVGET-KQAPMELE 1343
 QY 1180 NKASP-DWPOGEVTEENEMERYRENPLVLKVSFTIKPERKIGIVRTSGSKSLGM 1238
 Db 1344 QDKKPKNWPQEGVEFQNFQVRYREGIDLVRGVSFNIOGGEKVGIVGRTGAKSSITL 1403
 QY 1239 ALPLVELSGGCIKIDGVRSIDIGLADLRKSLTIPOEPVLFSTGVSNIDPFNOYEDQ 1298
 Db 1404 ALFRIILIAAGRISIDIVDIAWSGLHMRSLRTIIPQDPVILSGSLRNIDPFKIDDE 1463
 QY 1299 IMDLEERTHMEKCIATQPLKLESEVMENGNFVSGEROLICITARALLRHCKIILIDEXA 1358
 Db 1464 IWRLELSHLKSPKSLAAGINHEIAGENGSLVGQROLVCLARALLRKTPVILVDEXTA 1523
 QY 1359 AMDTEITLLIQTITREAFADCTMLTIAHRLHTVLGSDPRINVLAAQOVVEFDPVLLSND 1418
 Db 1524 AVDETFDLDLQ-----FKECTVLTIAHRLNTIIDSDKVIYLDKGQILIEFASPTEILDNP 1577
 QY 1419 SSREYAM 1425
 Db 1578 KSAPYSM 1584
 RESULT 9
 ID 024635 PRELIMINARY; PRT; 1622 AA.
 AC 024635;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
 DE MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN HOMOLOG.
 GN MRP1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI-Taxid=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=97368351; Pubmed=9223346;
 RA Lu Y.-P., Li Z.-S., Rea P.A.;
 RT "MRP1 gene of Arabidopsis encodes a glutathione S-conjugate pump:
 RT isolation and functional definition of a plant ATP-binding cassette
 RT transporter gene.";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:8243-8248(1997).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Lu Y.-P., Li Z.-S., Rea P.A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Lu Y.-P., Li Z.-S., Rea P.A.;
 RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 TRANSPORTERS).

DR EMBL: AF008125; AAB71832.1; -
 DR EMBL: AF008124; AAB67319.1; -
 DR HSSP: P13563; INBD.
 DR Mendel: 24381; Arabid.1944;24381.
 DR InterPro: IPR001140; -
 DR InterPro: IPR001687; -
 DR InterPro: IPR003439; -
 DR InterPro: IPR003593; -
 DR Pfam: PF00005; ABC_tran. 2.
 DR Pfam: PF00664; ABC_membrane. 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER. 2.
 DR SMART: SM00382; AAA. 1.
 KM APP-binding; Hydrolase; Transport.
 SQ SEQUENCE 1622 AA; 182015 MW; 6f0287af651ad5cf CRC64;
 Query Match 27.6%; Score 2013.5; DB 10; Length 1622;
 Best Local Similarity 34.3%; Pred. No. 3.9e-128;
 Matches 476; Conservative 240; Mismatches 501; Indels 171; Gaps 21;
 QY 74 EEPKGYHHGLSALKPIRTTCKHQHVDNAGLFSCMTFSLSLARVAHKGELSMEDV 133
 Db 217 EELPGGE-----NICEPRH---ANLPDSIFFSWLNLMTLGSKR-PLTEKDV 259
 QY 134 WSLKHSDDVNCRLRLMOEELNEVGPDAASLRVYWIICRTRLILSTICLMTQLAG 193
 Db 260 WHLDWTMDKTEIHLMSFOQSWDELEKRP-----W-----LIRALNLSLG 300
 QY 194 -----FSGPAFWKHLELYQATES---NLQYSLLVGLLLEIVRSW 234
 Db 301 RFWMGGEFKIGNDGSGFVGP-TLNLLELKSQMLNEPAMIGIYVLSFVGVLGICE-- 357
 QY 235 SLATMALNTGTGYRLGAILITMAFKILKLNKESSL--GELINICSDNGQMFEAA 292
 Db 358 --AQYFQVMVNGYRLSALTAFAFRKSLRLTNEGRRKFKQGTNTMTTAEISLQIQ 415
 QY 293 VGSLLAGPVAIIIGMIVNIIILGPTGLGSAVFLLFPAMMFASRIFAVPRRCVATD 352
 Db 416 SLHTWMSAPRIITVALVILYQOLGVAISIGLFLVLFPIQTVIISTOKITKGIDRTD 475
 QY 353 ERVOKMEVLTYYIKIKMYAVAKASQSVOKIREERIRLEKAGYFOSITGVAPIVVI 412
 Db 476 KRIGLMNEVLAAIMTVKCYAMENSFOKVTGRDDELSPFKAKLLSAFNNFILNSIPVL 535
 QY 413 ASVTFSVYHMTLGFDLTAQAFVYVYVFNSTFPAIKYTPESVKSLSASVAVDRKSLFL 472
 Db 536 VTVVSFGVSLGDLDPARAFSTLSLFSVLRFPFLMIPNIIITQMVANVSLNLEEVLS 595
 QY 473 MEYHMIKKNKASP-HIKIEKNATLAWDSHSSIONSPLKTPMKKDKRASRGKKEVR 531
 Db 596 TEERVLLPNPIEPGQPAISIRNGYFSWDS----- 625
 QY 532 QLOTEHOAVLAEQGHLLDSDERSPEEKEGKHILGHLRLQTLHSIDLETOEGKLY 591
 Db 626 -----KADRP-----TSLNIMLDIPLSLV 645
 QY 592 GIGSGVSGKTSLSAILIGM-TLEGSIAISGTPAYAAQOAMLNATLNPILFGKEYD 650
 Db 646 AVVSGTSGKTSLSIAMGELPARSDATVLRGSAVAYPOVSWTFNAVARNILLEGAFD 705
 QY 651 EERNYSVNSCCLRPDLALPSSDTEIGERGAMNSGGORISLARLYSDRSYIIDD 710
 Db 706 QEKERVYIDVYALDHDELPLGGDLTEIGERGVNISGGQKORVSMARVYNSNSDYLDE 765
 QY 711 PLSALDAVGNHIFNSAIRKHLKSTVLFVHNOYLVDCEVIFMKKGCITTEGTHEEL 770
 Db 766 PLSALDAVGOVEKCKIKRELQGTTRVLTNOLHFLSQVDKILLVHGCVKEGSTYEEEL 825
 QY 771 MNLNDVATITNNL-----LIGETPPVEINSKKETSGSQKSODKPGPTGSIKKEKA 822
 Db 826 CHSGPLFRLEMNNGKVEDYSEENGEAEVHOTSVKPVENGANNILQKGIITKSKSENS 885
 QY 823 VKPEGQLVQLEKGGGGSVPWSVGYVTAAGGGLAFVLVIALPFLNLNGVAFSTWMLSY 882

```

Db 886 V-----LVKREERETGVYSMKVLERQNALGAMVYVMLVICYVLTQYFRVSSITWLS 939
Qy 883 WIKOSGNTTARGNTSVSDSKNDPHMOYASITALSMAVLLIKAIRGVVFGTAR 942
Db 940 WIDSG---TPKHG-----PLFYNIYALLSFGVSVTLINSYWLINSSLY 982
Qy 943 ASSRLDELFRRLSRPMKFPDTPPTGRILNFRSKMDVDVLPQAEFIONVILVFE 1002
Db 983 AKKMHADALGSLRPMVFPOTNPLGRILNFRPAKMGIDRVAVFVNMFGSIQOLLS 1042
Qy 1003 CVGMIAGVPEWELVAVGPLVILFVSLVILSVRLRELKRLDNTQSPFLSHITSIQGLA 1062
Db 1043 FYLLIGIVSTLSLMAIMPLLVFVGYALTYQNTSRREIKRMDSTTRSPVYAQFGEALNGLS 1102
Qy 1063 THAYNKGEPFLHRRQELLDQNAPEFLFTCAMRLAVLDLISLITTTGMLVILMG 1122
Db 1103 STRAKVADMAEINGRSMDNNIRFTLVNMAANRMVIGILEVIGLVMWTASLAVMONG 1162
Qy 1123 QT--PRAVA---GLAISYAVOLTGLFQFTVRLASEFEARTSVERINHVIKTLSEAPAR 1177
Db 1163 KANQQAIVSTMGILLISTLSTSLTAVRLASLAENSLNSYERGVNTEIPS-BAPLV 1221
Qy 1178 IKNKAPSPDMPQEGEYTFENAEKRYRENLPVLYKKVFTIKPEKIGIVRTGSGKSLG 1237
Db 1222 IENNRPPGMPSSGSIKFDVYLARKYRELPLVYLGVSFLISPMKDVIGYIRGTAGKSSLL 1281
Qy 1238 MALFRLVELSGGCIKIDGVRISDIGIADIRSKLSTIPQEPVLFSGVTRSNLDPFNQYTD 1297
Db 1282 MALFRLVELSKRILDECDIRFGLMDIRKVVGLIPQAPVLFSGVTRSNLDPFSEHNDP 1341
Qy 1298 QIMDALERTHMEKCTAOLPLKLESEVMENGDNPSVGEROLLCAARALLHCKLITIDEAT 1357
Db 1342 DMESIERHLDOTIRKNPGLDAEYTAGENFSGORULLSLARLKRSLVLIDEAT 1401
Qy 1358 AAMDETDLIOETREAFDCTMLTIAHRLHTVLSGDRIMVLAQOVVEFDTPSVLLSN 1417
Db 1402 AAVDVTDLIOKTRIEERKSCMLIAHRLNITIDCKRVILDSKQVDFESSPENLISN 1461
Qy 1418 DSSREYAM 1425
Db 1462 GESSFSKM 1469

```

```

RESULT 10
Q9N2N3 PRELIMINARY; PRT; 1528 AA.
AC Q9N2N3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
DE C. ELEBANS MOLTIDRUG RESISTANCE RELATED PROTEIN 1 (MRP-1) (GB:066260),
CONTAINS SIMILARITY TO PFAM DOMAINS PF00664 (ABC_MMBRANE,
E-VALUE=1.93e-5, E-VALUE=1.1E-54, N=2) AND PF00005 (ABC_TRAN, SCORE=277.3,
DE MRP-1.
GN MRP-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA "Genome sequence of the nematode C. elegans: a platform for
RA investigating biology. The C. elegans Sequencing Consortium.";
RT Science 282:2012-2018(1998).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A.;
RT "The sequence of C. elegans cosmid F57C12.";

```

```

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
DR EMBL: 041554; A031550.2; -
DR InterPro: IPR001140; -
DR InterPro: IPR001687; -
DR InterPro: IPR003439; -
DR InterPro: IPR003593; -
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00664; ABC_membrane; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR SMART: SM00382; AAA; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 1528 AA; 171046 MW; 3CFC5D831C0EA0A5 CRC64;

```

```

Query Match 27.5%; Score 2013; DB 5; Length 1528;
Best Local Similarity 33.9%; Pred. No. 3,8e-128;
Matches 488; Conservative 239; Mismatches 483; Indels 228; Gaps 28;

```

```

Qy 96 KQHPPVNDGLPSCMFPMSLARSVAHKKGELSMEDVMSLSHSSDVNCRRLRLMOE 155
Db 204 KNSCPETYSFLNQLTFFQMFSGIAYIGNKK-SLEKEDLDLNRD-----KAENIIPS 255
Qy 156 ELNEVSPDASLR-----VWIFCRTL-ILSTVCL-MITOLAFSG 196
Db 256 FIENLIPVEGYRKRIRKNPEAIRKNHSILIPITKTKFTLLAGCKYKIMDLQFYA 315
Qy 197 PAVWAKHLEIYQATESNLQYSLVGLLLEIYVSWSLATWALNRYTVARCAIIT 256
Db 316 PE-ILQLISFIEDKQNPWIGVSIALLMFLSSILQSMILHGFHMFRLGMNIRSLVS 374
Qy 257 MAFKRLIKLKN--IKKSLGELINICSDNGORFEAAVAGSLAGPVAAILGMITNVII 314
Db 375 AVYTKTLNLSNEARKGTGAIYNLMSVDIQRIDQMTPTIMEFWSAPILILSLYFLMKL 434
Qy 315 LGPTGFLGSAVFLFYPPAMFASRLTAFFRRKCVATDERVOKMNVLYIRKIKYAMV 374
Db 435 LGVSVLAGFVILLILIPMSFISVKMRNCOMQMKRDERIKMSILNGMKYLKYSME 494
Qy 375 KAFSGSVOKIRIEERRILKAGYFOSITY--GVADIVVIVASVTVFSVHMTLGFDP--L 428
Db 495 KSMKKVLEVREREIRVLLKLSYLNATLISWACAPFLV---AVLTFGLYVLMDPENNVIL 551
Qy 429 TAAQATVTVFNSMTFALKVYTFESKSLSEASVAVDRKSLFLMEEVHMKRKPSPH 488
Db 552 TPQITVALALFELIFPLAVFAMFSAOVCSASWTRKEFFAAEM-----SPOT 603
Qy 489 K-----IEKNNTLMDSSHSISQNSPKLTPMKKDKRASRCKKEKKEVROLQRTED 539
Db 604 STAYGTDASAKMDGGSFAM-----GSKEE----- 628
Qy 540 AVLAQKGHLLDSDRPSPEEERKHIHLGLRLQRTLSIDLETQELKVGICSGVGS 599
Db 629 -----DRKLHDTIFNIKROLVALVAVRGVS 653
Qy 600 KTSLSIALIGQMTLLGSAISGTFAYVAQAMITNATLRDNIILGKEYDERKYSVLN 659
Db 654 GKSSLLHALIGEMNKLKSGSVQVGSVAYVQOLAMIONLSLRNNILNRPDALYQNVIE 713
Qy 660 SCCLRPDLALIPSSDTEIGERCANLSGGRORISLARALYSRSYIILDDPRLADAHY 719
Db 714 NCALYODLESIPREDTEIGEKINLSGGOKQVSLARAVYQNAELVLLDDPRLASDSHY 773
Qy 720 GNHIFNSAIRKH--LKSRTVLEVTIHOLOYLDCEVIFMKEGCITERTGTHEELANLND 776
Db 774 GKHIENVISIATGCTGTATVLLTGLTYLKHCDQVIVYLKDETIEMGTYQELMNSNGA 833

```


QY 884 -----IKSGSNTVTYRGNETSVSDSKMNDPHMOYASITVALSMAVMLLKAINGVVFV 937
Db 1010 DDAKIALISGMSSET-----OIRLGITAVLGMGATSCAASITMA 1052
QY 938 KGTLRASSRLHDELFRRLILRSPMKFFDTPTGRLINRFSKDMDEVVRLPFOAEMFIONV 997
Db 1053 LGMVCAASHLHNTLEENRSMFAFDVTPGLRLINRFGKDVGDVTDIPRMSFIRTA 1112
QY 998 ILVFFCVMGIVGFVFWFLVAVGPVLVLSVLAHSRVILRELKRLDNITQSPFLSHITSS 1057
Db 1113 VSSILRIIILIMAPPLAISLLPLPIRFYV-----STRQIKRLESASRSPISYHFQES 1166
QY 1058 IGGATTAHYKNGOELRHYOEILDNDNOAPFLPTCAMRWLAVRDLI-STALITTTGLM 1116
Db 1167 IGGASSIRAGVYDKFIESQHRVDENLATIYPSIYANRWLAVRLEMGNLVLVSAGAA 1226
QY 1117 IYLMHQ-IPPAVAGLAISYAVOLTGLEFQFVRLASETEARPTSEVERINHYIKTLSLEAP 1175
Db 1227 VYFRDSPGLSAGLVGLSYVALNITQTLNMAVRMTSELETINVAVERINEX--TTPREG 1284
QY 1176 ARIKKASPDMPDGEVTFENAEKRYRENLPVLTKVYSFTTKPEKIGIYGRSGKSS 1235
Db 1285 NNSOGLAEK-SWPEGEISIKNFVYRPGDLVLHGYTAHISPECKIGIYGRSGKSS 1343
QY 1236 LGMALFRLVELSGGCIKIDGVARISDGLADRSKLSITPOEFLPSGTVRSNLDPEFNOYT 1295
Db 1344 LTLALFRILEADGGCIEIDGTIADILLEQLRSRLTIYPODVLSGVRKRMMLDPPAFS 1403
QY 1296 EDQIDALERTHMKECIALQPLKLESEVMENDNPSVGERQLLCLARALLRCKTILIDE 1355
Db 1404 DDQIWEALRNHLDSEFVKSLOGELHHHISEGENDLSVGROQLICARALLRCKTILIDE 1463
QY 1356 AFAANDETDLIIQITIEAPDCMTLITANRLHVLVLSDRIMVLAQGVYFEDIPSVLL 1415
Db 1464 AAAAVVETDLSLQGTIRQEKDCVTLTIAHRLNTVMDSDRLVLIDKGVAFEDIPKVL 1523
QY 1416 SNDSRFYAM 1425
Db 1524 SNPDGIFYSM 1533

RESULT 12
ID 048907 PRELIMINARY: PRT: 1623 AA.
AC 048907:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE MGATP-ENERGIZED GLUTATHIONE S-CONJUGATE PUMP (EC 3.6.1.3).
GN MRP2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lu Y.-P., Li Z.-S., Drozdowicz Y.M., Hortensteiner S., Martinola E.,
RA Rea P.A.;
RC Plant Cell 10:1-18(1998).
CC -I- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
CC EMBL: AF020288; AAC04245.1; -.
DR HSSP: P13569; INBD.
DR Mendel: 24660; Arabid.1944;24660.
DR InterPro: IPR001140; -.
DR InterPro: IPR001687; -.
DR InterPro: IPR003439; -.
DR InterPro: IPR003593; -.
DR Pfam: PF00005; ABC_tran. 2.
DR Pfam: PF00664; ABC_membrane. 2.
DR PROSITE: PS00211; ABC_TRANSPORTER. 2.
DR SMART: SM00382; AAA; 1.

KW ATP-binding; Hydrolyase; Transport.
SQ SEQUENCE 1623 AA; 182072 MW; 9630B7B04899970 CRC64;
Query Match 27.3%; Score 1993; DB 10; Length 1623;
Best Local Similarity 34.7%; Pred. No. 9.8e-127;
Matches 481; Conservative 230; Mismatches 483; Indels 192; Gaps 29;
QY 95 CKHQHPVDNAGLFCSCMTWSLSTARVAHKKGELSMEDVWSLSKHESSDYNCRRLERLMQ 154
Db 226 CPEKHP-----NIFDKIFFSWMNPLMTLSKR-PLTEKDVYLDTDQDETLETSPQSHSD 280
QY 155 BELNEVGPDAALRRVNVNFCRTLRILSLIVCLMTIQLAG-----FSGPA 198
Db 281 KELQKPP-----W-----LRLANSLSGRFMGGFWKIGNCSQFVGLP 321
QY 199 F---WVKHILEYQATESNL-QYSLL--LVGLLLETVRSMSLALTYALNRYGVRLNG 252
Db 322 LNLQGLSKMQEDAPAMMGITYAFSIFGGVFGVLC-----AQYFQNVKRGVRLNS 373
QY 253 ALLTMAFKKILKAIKIKESL--GELINICSDGQRMFEAAVGSLLAGCPVALIGMTY 310
Db 374 ALIAAVFRKSLRTNEGRRKFPQTKITNMTDAESLQICSLHTMMSAPRIITALL 433
QY 311 NVTLIGPGFLGSAVFILFEPAMMFASRLTAYERRKCVATDERYOKMNEVLTYSIKFM 370
Db 434 LYQGLVASLIGALLIVLMFPLQTYIISKQKLTKEGLQRTTKRIGLMEVLAAMDYKVC 493
QY 371 YAVMAFSGSVOKIREEBRRILEKAGYFOSITVGVAVPIVYIASVYTESVHNTLGFDLTA 430
Db 494 YAMENSFGSKQVTVRDELDLSWRKSQLGALNMFILNSIPVLVTIYSCFVPLLGDLTP 553
QY 431 AQAFTVYVFNKNTALAKTTPRSVKSLSASAIVDRFKSLFMEVYHMKKNRPAAP-HIK 489
Db 554 ARAFTSLSLFAVLRPLFLPNIITQVNAVANSLNKLEEVLAETEEKILLPNPPIEGEPA 613
QY 490 IEMKNATLAMDSSHSSIONSPLTPPKMKDKRASRGKKEKVQLORTEHQVLAQKGLH 549
Db 614 ISTRNGYFSWD-----KGS--- 627
QY 550 LLDSDRPSPEEKEGKHNLGLRLQRLTSLIDLEIOEGKLVGSGVSGKTSLSIALI 609
Db 628 -----DRP-----TLSNINLDVPLGSIVAAGVSGTEGTSLSIALI 663
QY 610 GOMTLLEGSI-AISGTPAVVAQAMILNATLPDNLIFGEYDEERYNSVLNSCCARPDLA 668
Db 664 GELPRTSDAIVTLRGSVAAYVQVSWTFENATVNDLIFGSPFPERKERAIIDVTSKHDLE 723
QY 669 ILPSSDLPTFIEGRGANLSCGORORISLARALYSDRSIYLLDDPLSLADHVGNIHFNSAI 728
Db 724 LPLGGDLTEIGRGVNISSGQKORVSMARAVSNSDVILFDDPLSLADHVGQOYFEKCI 783
QY 729 RKLHLSKTYLFTYTHOLQYLVDCDEVLFMKEGCITERGTHEELMNLNGDYATLFNNLLIG- 787
Db 784 KRELQKTRVLVTLNDHLFSLQYDRIVLVHEGTVKEGTYEELSS--NG---PLFQRLMENA 839
QY 788 -----ETPPEVINSKKE-----TSGSQK--SQDKPRTGSIKKKAVKPEEGQLV 831
Db 840 GRYEYSESENGEAEADQTEQPVANGNTGLQMDGSDDKKSEKNGKGSV-----LI 893
QY 832 QLEEKGGGVSPSYGYVYQAAGGPLAFIVIALFPLANNGSAFSTWMLSYVIKQSSGT 891
Db 894 KOBERETGVASRWVLLKRYODALGAMVVMMLLCYLLLEVFVYTSSTWLSSETWDAG--T 950
QY 892 TYTRGNETSVDSKMDPNMHOYASITVALSMAVMLLKAIRGVFVFKGTLRASSRLHDEL 951
Db 951 PKSHG-----PLFYNLITALLSFGQYLVLTITNSYWLIMSLVYAAKKIHDMA 996
QY 952 FRRLILRSPMKFFDTPTGRLINRFSKDMDEVVRLPFOAEMFIONVILVFCVGMATGVF 1011
Db 997 LHSILRAPSPFHTNPLGRILINRFAKDLGIDIRTVAVFVNMFGVQSLSLSTVILGIYS 1056
QY 1012 PMFLVAVGVLVLSVLAHSRVILRELKRLDNITQSPFLSHITSSIGLALITIAHYNKQ 1071

```

Db 1057 TSLMAMPLPLVLYGAYLYQNTAREVKRMDSTISRSPPVYAQGFGEALNGSTIRAYK--- 1113
QY 1072 EFLHRYOELDD-----DNQAPFLEFTC-AMRWLAVRDLISALITTTGLMIVLMHQI 1124
Db 1114 ----AYRMADINRSMDNNRFLVVMGANKRWIGIRLETGLMIMWLTASFVAMQNGRA 1169
QY 1125 --PPAYA---GLAISVAVOLTGLFQFTVRLASETEAPRTSYERINHYIKTSLDEAPARIK 1179
Db 1170 ENQAPFASTWGLISVALINTSLTGLVRLASLANSINLAVERGVNIE-IPPAAPVPIE 1228
QY 1180 NKAPSPMPQGEVTEFNAERRENTPLVLYKYSFTIKPEKIGIYGRSGSSIGMA 1239
Db 1229 NMRPPGMPSSGSIKEFDVYVLRPQLPVHGVSEFTIHPDKGIVGRGAGKSSLLA 1288
QY 1240 LFRVLSEGGCIRKIDGRISIDIGLADLRSLKSLIPOEVLFGSTVRSLDPFNQYTEDQI 1299
Db 1289 LFRIVEVEGRLLIDDCODVGFGLMDLRKVLGIIQSPVLFSGVFRFLDPFGHNDADL 1348
QY 1300 WDALERHMECTIAQLPLKLESEVWENGDNFSVGEROLLCAIRALLRHCKILLDEATAA 1359
Db 1349 WESLERAHLDKTIERNPLGLDAEVSAGENSFGOROLLISRALLRSKILLVDEATAA 1408
QY 1360 MDRETDLLIOETIREAPADCTMLTAHRLHTVLSGSDRIMVLAOGVVEPDPSPVLSNDS 1419
Db 1409 VDVTDLIOETIREAPADCTMLTAHRLHTVLSGSDRIMVLAOGVVEPDPSPVLSNDS 1468
QY 1420 SRFYAM 1425
Db 1469 SSFSKM 1474

RESULT 13
064590 PRELIMINARY; PRT; 1623 AA.
AC 064590:
DT 01-ANG-1998 (Tremblrel. 07, Created)
DT 01-ANG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN, ATMRP2.
GN T29F13.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_taxid=3702;
OC NCBI_taxid=3702;
RN SEQUENCE FROM N.A.
RP STRAIN-CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.,
CC Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
DR EMBL: AC003096: AAC16268.1; -.
DR HSP: P13569: INBD.
DR Mendel; 29399; Arabid; 1944; 29399.
DR InterPro: IPR001140; -.
DR InterPro: IPR001687; -.
DR InterPro: IPR003439; -.
DR InterPro: IPR003593; -.
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00664; ABC_membrane; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR SMART: SM00382; AAA; 1.
KM ATP-binding; Transport.
SQ SEQUENCE 1623 AA; 182128 MW; F074F0F6ED7A4D47 CRC64;

```

Query Match 27.3%; Score 1992; DB 10; Length 1623;
 Best Local Similarity 34.7%; Pred. No. 1.le-126;
 Matches 478; Conservative 231; Mismatches 487; Indels 182; Gaps 27;

```

QY 100 PVNAGLFSCFTSFSSLSLAVNAHKGELSMEDVYLSKHESSVNCRRLERLMOEELNE 159
Db 227 PEKANIEFDKIFFGSMWMDPLMTIGSKR--PLTEKDVWYIDTMDOTETLFTSFQHSMDKELOK 285
QY 160 VGPDAASLRVWVIFCRRLLSIVCLMITQLA-----SSGPAF---M 200
Db 286 PGP-----W-----LIRALNNSLGRFWMGCGWKIGNCSQFVGLLNLQL 326
QY 201 VKHILEYQATNESMLQYSLVLGLDLLEIYRSWSLALTYALNRYTVRGALITLMAFK 260
Db 327 LKSMQEDAPAMNGYT-YAFSIFGVVEGVLC-----AQYQNVAVRGVRLSALIAAVER 381
QY 261 KILKLNKEKSL--GELINICSDGQRMFEAAVGSILAGPPVAILGMVNIILGPT 318
Db 382 KSLRLTEGRRKFTQGTIKTMTMTDASLQICOSLHTMWSAPRIIALLILYQOGLVA 441
QY 319 GFLGSAVILFFYPAMPFASRLTAVFRKCVATDERVQKMEVLYIKFIKIVNAWAFS 378
Db 442 SLIGALLVLMFPLQTYIISKQKLTKEGLORTDKRIGLMEVLAAMDYVCYAMENSFO 501
QY 379 QSVOKIREERRILEKAGFPOSITGVAPIVVYASVYTESVHNTLGFDTLTAQAFTVT 438
Db 502 SKVQTVRNDDELMSWRKSQLGLALMFLINSIPVLVTIYSEGVFTLLGGDLTPARAFSTLS 561
QY 439 VENSMTALKVTPPSVKSLSBASVAVDRFKSLFLMEEYHMKRPPAS-RIKIEKNAITL 497
Db 562 LFAVLRPLFPLMIPNIIQVAVANVSLKRLLEVLTATEERILLPNPPIEGEPALISIRNGYF 621
QY 498 AMDSHSSIQNSPKLTPFKMKDKRASRGKKKRVQLORTHEQVLAQKGLLDSDRP 557
Db 622 SWDS-----KG-----DRP 630
QY 558 SPEDEEGKHIMGLRLQRLHSIDLEIOEGKLVGIGSVSGKTSLSIALIGQMTLLG 617
Db 631 -----TLSNINLDVPLGSLVAVVSGTEKTSLSIALIGELPATSD 671
QY 618 SI-AISGFAVVAQANILNATLNDNILFGEYDEKRYNSVLNCCRLPLALIPSSDLT 676
Db 672 AIVTLRSVAAPVQVSWTFENATVRDNLIFGSPFREYTERAIDVTSIKHDELLPGDDL 731
QY 677 EIGRGANLSCGORISIALRALSIRSISYITLDPLSALDAVNHIFSAIRKHLKST 736
Db 732 EIGRGVNISSGQKQKRAMAVYNSDVYIFDDPLSALDAHVQOYFEKCIKRELQKT 791
QY 737 VLEVTHQLYVDCDEVYFMKEGCTERGTHEELMNLNGDIATIFNNLLG----- 787
Db 792 RVLVTNQLHFLSQVDRIVLYHEGTVEKGYEELSS--NG---PLFORLMEVAGKVEEYSE 847
QY 788 ETPPEVLEINSKKE-----TSGSOKK--SODKGPRTGSIKKRKAYKPEGGQYUOLEEGOG 839
Db 848 ENGEAEADQTAQPVANGNTNGLOMDSDDKSKKGGKAGSV-----LTKOEEKRG 901
QY 840 SVPSWVGVIYIOAGGLAFLVIALMPLNANGSTAFSTWMLSYIKOGSGNTYTRGNET 899
Db 902 VVSWRVLRKYQDALGAMVWMLLCLVLEFVEFVTSWTLSMTDAG---TPRSHG--- 955
QY 900 SVSDSMKDNPMQYASIALSAVMLLKAIRGVVYVYKGTLRASSRLHDELFRRIIRSP 959
Db 956 -----PLFYNIYALISFGOVLVLTNSYWMISLSYLAARKLHMMLSHIRAP 1004
QY 960 MKFPDTPPTGRLINRFGSDMDEVVRLPFOAEMFTQWVILFFCVGMIAGVFPFLVAVG 1019
Db 1005 MSFHTNPLGRILINRFKADLDDIDRTYAVFYVNMGMGOVSOLSTVYLIGIYSTLSLWAIM 1064
QY 1020 PLVILFVLIHVSRLIRLEKRLDNITQSPFLSHITSSIOGLATIHAYNKGQEFILHYOE 1079
Db 1065 PLVILFVGIYLYNTAREVKRMDSTISRSPPVYAQGFGEALNGSTIRAYK-----AYDR 1117
QY 1080 LLD-----DNQAPFLEFTC-AMRWLAVRDLISALITTTGLMIVLMHQI--PPAYA- 1129
Db 1118 MADINGRSMNNIRFTLVVMGANKRWIGIRLETGLMIMWLTASFVAMQNGRAENQAFAS 1177

```



```

Db 1366 VPDPVFLFSGMKMKMLDPEFSASVSSQSVAEALLENALMKRFVSLDQGLKHKISKGGENISLV 1425
OY 1333 GEROLLCIARALLRCKILLIDEATAADTETDILLIQTETREAFADCTMLTAHRLHVL 1392
Db 1426 GOROLICILARLLKRLTKVLVEDEAAAAADVDSIDTQITREOFECTVLTATHRLNTVM 1485
OY 1393 GSDRLMVLVAGGVVEEDRPSVLSSNDSSRYAM 1425
Db 1486 DSDRLVLVDKGRVAEEDSPKMLANPDGIFYSM 1518

RESULT 15
O20943
ID O20943 PRELIMINARY; PRT: 1494 AA.
AC Q20943;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SIMILAR TO THE HUMAN MULTIDRUG RESISTANCE-ASSOCIATED PROTEIN.
GN F57C12.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelioderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jler M., Johnston L.,
RA Jones M., Keshav J., Kirsten J., Laister N., Latelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Straden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkins-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX Favello T.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX Waterston R.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC -1 SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
CC TRANSPORTERS).
CC EMBL: U41554; AAA83299.1; -.
DR HSSP: P13569; INBD.
DR InterPro: IPR001005; -.
DR InterPro: IPR001140; -.
DR InterPro: IPR001687; -.
DR InterPro: IPR003439; -.
DR InterPro: IPR003593; -.
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00664; ABC_membrane; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR PROSITE: PS00037; MIB-1; UNKNOWN_1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding; Transport.
SEQUENCE 1494 AA; 166742 MW; A88A6EC8F683AFDC CRC64;

Query Match 27.2%; Score 1984.5; DB 5; Length 1494;
Best Local Similarity 34.5%; Pred. No. 3.2e-156;
Matches 493; Conservative 232; Mismatches 499; Indels 203; Gaps 31.

```

| | | | |
|----|------|--|------|
| QY | 152 | LMQELNEVGPDAASLR-----VWLF--CRTRLISLYCMLITQDLAGSGPAFVNRLL | 205 |
| Db | 222 | YHOKIKRD--PSALPKPNHSPFVPIFKYTKYTLLAGFFYKCLCEMDLPQLAQ--LTKOLI | 278 |
| QY | 206 | EYQATSESNLOYSLILVJLGLLEITEYRSWSLALTWALNRTGVRLRGAILMAEFKILKL | 265 |
| Db | 279 | GFIEDKNQPMWIGOSLYGMFSSFOJSMLEHQYHSMRKLAMHVRSVLTSAAVYSKALNL | 338 |
| QY | 266 | KN--TKRSJLGEILNICSNDGQMFZAAVGSLLAGGPVVALIGMITYNIIIGPFGIGS | 323 |
| Db | 339 | SNEARCKTIGAIYNLMNSVDIOKIDQMAPTIMLFWSAPLQIFLSTYFMKEFGAVALAGL | 398 |
| QY | 324 | AVFLFPAMMFASRLTAYPRRC-----VAATDERQKNEVLTYIKFKIMYAWAKFSQ | 379 |
| Db | 399 | VLLILAP-----VNGLAIQMRKCQJDEQMKLADERIKMMSSELINGKYLKXSWRSMEN | 454 |
| QY | 380 | SVOKIREERRILIEKAGYFOSITVGVAPIVYVYIASVYFESVHMTLGFQ--LFAAQFTV | 436 |
| Db | 455 | MYLIRERRELIHLIKLSYFMAAIVFWMICAPLASYISFVVYVYLDPENNVLTPEITFVA | 514 |
| QY | 437 | VTVNSMTFPAKTYPPSVKLSASAVAVDRKSLFLMEVHIKKKKPASPPIKLEMKAT | 496 |
| Db | 515 | LISLEDLIRMPILAMVAMVYGAOVOSVSNRLEEFPAEEM-----SPQSTI----- | 560 |
| QY | 497 | LAMDSHSSSIQNSPKLTPKKKKKKRASRGKKEKRVQLOTEQOAVLAEQGHLILDSDER | 556 |
| Db | 561 | -----SHG-----EIDSAIIEVENGFLSMSDED | 583 |
| QY | 557 | PSPEEBGKHIIHGLRLQTLRLSHIDLEFQEGKLVGCGSVGSGTSLISAILQMTLLE | 616 |
| Db | 584 | P-----TLREISFKIQOGVOLAIVKVSQSGSSLHILHGLGEMNLS | 624 |
| QY | 617 | GSIAISGTFYVAQOAMILNATLRDNLIEGKEDEERYKSVNSCCPLDAILPSSDT | 676 |
| Db | 625 | GSVQINGNTIAYVQOAMIQMSLRNNIILNRKPIDELENTDYKKNALKEDLANLPAGRT | 684 |
| QY | 677 | EIGERGANLSQGORISITARALYSDBRSYIILDDPLSALDAHGVNHIFNSAIRKH--LK | 733 |
| Db | 685 | EIGKRGINTLGGGQKQRYSLARAYQONPDITLLDDPLSADSVHGKHIFENYISSGTCLA | 744 |
| QY | 734 | SKVLPFTHOLOLVYDDEVETPEKKECITERGHEELMLNDGVATFENNLL----- | 785 |
| Db | 745 | SKTVLVYTHGLTYLKHCDQDLVYKEGTISELGTQYDGLLNNASAPFLEELFLESKTRG | 804 |
| QY | 786 | -----LGEPP-----VEINSKKTSG-----SOKKS-- | 807 |
| Db | 805 | RVASIGDGSGBEILDLQGVKPGILKRLSEHSQSEKEDTSARALIEYSDSSRRSVL | 864 |
| QY | 808 | -----QDKGRTGSIRKEKAVKREBEOVLQLEKGGOSVPMGVYGVYIQOAGPLAF | 859 |
| Db | 865 | LHSPRSQHEHEBALLGAISEDDPAQ--ENTQOLLEKETVETGKAKFEVYIAYFPAISIPTL | 923 |
| QY | 860 | LVIALLFMLNLSTAFTWMLSYWIMXQSGGNTVTGNETSYSDSMKDNPMHOYASITYA | 919 |
| Db | 924 | L-----PFPIYVSGSGIGILSNFYLAFL--SDHAKSGNTSSDAKME-----LGITYA | 968 |
| QY | 920 | LNSAVMLILKAIKGVYFVKCTLAASSRLHDELFRRLISPKMFEPTTTPGRLIRFSQDM | 979 |
| Db | 969 | VLCMGOSFVYLLASIIITLGTVLRSRILHAGLGLGNTMRSPMAFFPYTPIGRLINIGDI | 1028 |
| QY | 980 | DEVYVRLPFOAMEIQVILVYFCVGMINGVPMFLVAGPLVILFSLVHLYSRVL--- | 1036 |
| Db | 1029 | EADIRLTPDYIRIMSKMIFENVAVTLVYIMMARPAGIA---FALLSVIYFIVLRYLISTS | 1085 |
| QY | 1037 | RELKRLDNITQSPFLSHITSSIOGLATIHAYNKGQEFELHRVQELLDNQADPFLETQCMR | 1096 |
| Db | 1086 | ROLKRLSEASRSPISYFQFOSIGASSIRAFGVNDNFIKOSQOARDVHDILIAVYSIVANR | 1145 |
| QY | 1097 | MLAVRDLIL--SIALITTTGMLVYLMHGQ--IPRAYGILASVAVOVLTLGFLQFVRLASPTE | 1154 |
| Db | 1146 | MLAVREMVGNVLVLSAAGAAVYFRDSPGISAGVLGSLVSTALNTQTLNMAVNRVTSLE | 1205 |

THIS PAGE BLANK (USPTO)